

**Induced Pluripotent Stem Cells (iPSCs)
for research and therapy:**

**induction of hepatic differentiation in iPSCs
and evaluation of their quality as a model of *in vivo*
development in the context of coagulation**

by

SARA CAXARIA

A thesis submitted for the degree of

Doctor of Philosophy of the University of London

ABSTRACT

Stem cells hold great promise for regenerative medicine as they have the potential to repair almost any tissue. The use of induced pluripotent stem cells (iPSC) offers several advantages over human embryonic stem cells (hESC). Nevertheless, an issue that has slowed the use of iPSC (as well as hESC) in the clinic is safety. The pluripotent capacity that gives these cells their regenerative potential also gives them tumorigenic abilities. Moreover, the reprogramming procedure of iPSC can also affect the quality and safety of the final population. In the search for safer iPSC, we optimized an integration free method of reprogramming that is GMP compliant, which represents a step closer to their clinical use.

Hepatocyte-like cells derived from iPSC have proven useful in research, as in disease modelling and toxicity screening, and in the context of cellular therapy could represent a breakthrough for the treatment of liver disorders. We worked on an optimized method that allows quick and efficient hepatocyte differentiation from iPSC. We used this as a model to tap into embryonic development in the context of coagulation, and into the regulators of coagulation, with the goal to better understand coagulation. Similar studies can then be used to study other pathways besides coagulation and help increase the knowledge on the liver and its many functions.

DECLARATION

I, Sara Caxaria, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Work done on Chapter 1 and 2 was carried out under Dr. Pollyanna's Goh supervision and in parallel with her own work. We were each responsible for reprogramming different iPSC lines.

All animal handling was done either by Dr. Cecilia Rosales or Dr. Marco Della Peruta. Mouse embryos dissection was performed by Dr. Ariane Chapgier. All histological analysis was performed by UCL IQPath histological services.

Methods section 2.14 (Bioinformatics) was written by departmental Bioinformatics expert Pavithra Rallapalli. This method section relates to work done by Dr. Pavithra Rallapalli and by Professor Willem Ouwehand's next generation sequencing expertise group at Cambridge University (particularly Dr. Mattia Frontini, Dr. Myrto Kostadima and Dr. Romina Petersen).

RNA sequencing analysis work in chapter 5 was performed by Professor Willem Ouwehand's group at Cambridge University. RNA sequencing analysis work in Chapter 7 was performed by Dr. Pavithra Rallapalli and Professor Willem Ouwehand's group at Cambridge University. Motif and network analysis were performed by Dr. Pavithra Rallapalli.

Supplementary information 9.3 was written by Dr. Pavithra Rallapalli with more technical description of the work done in chapter 7.

Signed

ACKNOWLEDGMENTS

I would like to extend thanks to the many people, in many countries, who so generously contributed to the work presented in this thesis.

Special mention goes to my supervisor, Amit Nathwani. My PhD has been an amazing experience and I thank Amit, not only for the academic support, but also for giving me this wonderful opportunity.

Similar, profound gratitude goes to Pollyanna Goh, who has been a truly dedicated mentor. I am particularly indebted for your guidance and patience in teaching me so much throughout this PhD, and for your continuous support. I am also hugely appreciative to Marco Della Peruta and Sajjida Jaffer for your guidance and help in so many aspects of my PhD.

A big thank you to everyone in the ACN group for all the work, stimulating discussions and all the fun we have had in the last four year. Particularly, to all current and past members of the iPSC group: Pollyanna Goh, Susi Arthold, Sajjida Jaffer, Ariane Chapgier and Mahnaz Abbasian. Special mention also goes to Cecilia Rosales, Marc Davies, Doyoung Lee and Satyen Gohil.

Thank you to all my friends for your support, patience and for being there for me during the past 4 years. Thank you to: Leyla Mekkaoui, Solange Moscosso, Evangelia Kokalaki, Mariana Vieira, Filipe Ramos, Joana Coelho, Catarina Orvalho, Carolina Macedo, Margarida Palmela, Inês Sousa, João Bispo, José Fonseca, among others. Obrigada!

Special reference to Alexandra Kareh, who has been at my side from day one. Thank you for your constant faith in me. For your encouragement, support, patience and love. Without you this wouldn't have been possible.

Finally, but by no means least, thanks goes to my family (Mãe, Pai, Joana, Sofia, Paula, Avó, João e André) for your unbelievable support and love during this PhD and in my life in general. You have made this possible. You are the most important people in my world and I dedicate this thesis to you.

TABLE OF CONTENTS

ABSTRACT.....	2
DECLARATION	3
ACKNOWLEDGMENTS	4
TABLE OF CONTENTS.....	5
LIST OF FIGURES	8
LIST OF TABLES	11
LIST OF SUPPLEMENTARY FIGURES	12
LIST OF SUPPLEMENTARY TABLES.....	13
ABBREVIATIONS	14
Chapter 1 Literature review	17
1.1 Stem cells	17
1.2 iPSC reprogramming.....	22
1.3 Applications of iPSC.....	31
1.4 Questions and challenges of the iPSC field	35
Chapter 2 Materials and methods	41
2.1 Cell culture	41
2.2 Cloning	43
2.3 Reprogramming.....	44
2.4 Pluripotency	46
2.5 RNA isolation, RT-PCR and qPCR	47
2.6 Phenotype analysis by flow cytometry.....	48
2.7 Immunostaining.....	49
2.8 Hepatic differentiation.....	49
2.9 Periodic Acid-Schiff (PAS) Stain for Glycogen	50

2.10	ELISA Analysis	50
2.11	Liver failure model	50
2.12	Isolation of mouse embryos.....	51
2.13	GEO datasets	51
2.14	Bioinformatics	52
2.15	siRNA	57
2.16	Western Blotting.....	58
2.17	Primers.....	59
Chapter 3	Integration free reprogramming method.....	62
3.1	Introduction	62
3.2	Viral delivery systems for reprogramming	66
3.3	Episomal reprograming	71
3.4	RNA reprogramming.....	76
3.5	Characterization of the different lines	80
3.6	Discussion	85
Chapter 4	GMP compliant reprogramming method	86
4.1	Introduction	86
4.2	GMP compliant method	94
4.3	What comes next?	100
Chapter 5	Hepatic differentiation of human iPSC.....	101
5.1	Introduction	101
5.2	Hepatic differentiation.....	106
5.3	Tumorigenicity of HLCs	126
Chapter 6	Study of coagulation during development	141
6.1	Introduction	141
6.2	Do HLCs express coagulation factors?	155

6.3	What is happening <i>in vivo</i> ?	159
6.4	HLC as a model to study coagulation	171
Chapter 7	Regulation of coagulation factors expression	174
7.1	Introduction	174
7.2	In the search for a master regulator	176
7.3	Does HNF4a regulate expression of coagulation factors?	186
7.4	Discussion	195
Chapter 8	Final Discussion	198
Chapter 9	Supplementary Information	205
9.1	Supplementary Figures	205
9.2	Supplementary Tables	219
9.3	Motif analysis	220
REFERENCES		222

LIST OF FIGURES

Figure 1 – Waddington’s model on epigenetic landscape	17
Figure 2 – Modification of Waddington’s epigenetic landscape showing reprogramming possibilities of differentiated cells	20
Figure 3- Reprogramming process.....	22
Figure 4 – Graph showing the relation between efficiency and safety of different reprogramming methods	28
Figure 5 – Applications of iPSC	32
Figure 6 - Overlapping quality between iPSC and ESC.	36
Figure 7 – One-to-one relationship between RNA fragments and reads.	53
Figure 8 - Time-series RNA collection for analysis	53
Figure 9 - Background of transcript quantification using MMSEQ and MMDIFF.....	54
Figure 10 - iPSC generation methods	64
Figure 11 –Transduction efficiency of IMR90 fibroblasts using different AAV serotypes	67
Figure 12 - Comparison of GFP transduction efficiency of fibroblasts using retrovirus and rAAV.....	68
Figure 13 - Transfection efficiencies of human fibroblasts.	72
Figure 14 - Schematic of the cloning of the miR302 cluster into the pCXLE hOct4-shp53 plasmid.....	73
Figure 15 - Reprogramming with Yamanaka’s, Thomson’s and miR302 vectors.	74
Figure 16 - Comparison of miR302 expression levels.....	75
Figure 17 – mRNA reprogramming of BJ fibroblast line.....	77
Figure 18 - Comparison of mRNA reprogramming of patient Vs BJ fibroblasts line.....	78
Figure 19 – Characterization of established iPSC lines derived from BJ fibroblasts using retrovirus, plasmids and RNA.	81
Figure 20 - Characterization of pluripotency potential of different lines	83
Figure 21 - Human pluripotent stem cell scorecard assay results.....	84
Figure 22 – Efficiency of reprogramming using episomal method in feeder Vs feeder-free conditions.....	95
Figure 23 - Assessment of the best xeno-free matrix and medium for iPSC reprogramming.	96
Figure 24 – Comparison of episomal reprogramming under different culture conditions	98

Figure 25 – Simplified schematic of route of differentiation from pluripotent state to hepatocyte	104
Figure 26 - Schematic of the hepatic differentiation protocol	106
Figure 27 – Endodermal marker expression.	108
Figure 28 - Endoderm specific markers present in cell after endoderm differentiation	109
Figure 29 - Gene expression during hepatic differentiation.	111
Figure 30 – Cell density at day 4 of the differentiation protocol from two different iPSC cell lines	113
Figure 31 - Expression of liver specific markers during hepatic differentiation protocol.....	115
Figure 32 - Gene expression during hepatic differentiation protocol: stage 2 and 3	116
Figure 33 – Functional characterization of HLCs.....	117
Figure 34 - H&E of mouse Livers..	120
Figure 35 – HLC engraftment experiment analyses.	121
Figure 36 - Tumour in liver failure rescue with HLC.....	123
Figure 37 – Schematic of mouse liver failure experiment	124
Figure 38 – Complexity of tumours.....	127
Figure 39 – Hallmarks of cancer.....	128
Figure 40 - Expression of surface markers assessed by FACS.....	130
Figure 41 – Teratoma formation with cells at different stages of hepatic differentiation.	137
Figure 42 – Original cascade model representation.....	143
Figure 43 – Classical model of coagulation.....	144
Figure 44 - Initiation step of cell based coagulation model.....	145
Figure 45 - Amplification step of cell based coagulation model	146
Figure 46 – Propagation step of cell based coagulation model	147
Figure 47 - Table of expression of haemostasis specific genes in mouse embryos.....	150
Figure 48 – Coagulation system in lamprey.	153
Figure 49- Coagulation factors gene expression data in RDP2 iPSC line hepatic differentiation.....	158
Figure 50 - Coagulation factors expression in human foetal samples.	162
Figure 51 – Sections of human embryo at Carnegie stage 15.....	164
Figure 52 – Coagulation factors expression in mouse embryonic liver samples in GEO dataset.	166

Figure 53 – Coagulation factors expression in mouse embryonic samples at different time points of development.....	169
Figure 54 – Hierarchical clustering of coagulation factors expression detected in GEO dataset of cells during hepatic differentiation protocol.....	172
Figure 55 – Gene regulatory region where transcription is initiated	174
Figure 56 – Mean average (MA) plot representation of differential expression of genes in the sample overall (using DESeq2).	177
Figure 57 –Quantitative measure of expression of the extrinsic and common coagulation genes at the different stages of the hepatic differentiation protocol and in liver control.....	179
Figure 58 – Bar chart showing the fold change in all the 14 coagulation genes.	180
Figure 59 – Gene network for coagulation genes.	183
Figure 60 – Gene network for coagulation genes.	184
Figure 61 – Coagulation gene expression in HUH7 and HepG2 cell lines.	189
Figure 62 – HNF4a Protein levels after siRNA.	190
Figure 63 – Coagulation genes expression in HNF4a knockdown.....	193

LIST OF TABLES

Table 1 - List of fibroblast cell lines used for reprogramming	41
Table 2 – List of human Primers.....	59
Table 3 - List of mouse Primers.....	61
Table 4 – Summary of advantages and disadvantages of different reprogramming method...	65
Table 5 - Adapted teratoma assay to assess tumorigenicity of cells during hepatic differentiation protocol	133
Table 6 - Components of blood coagulation.....	141
Table 7 – Table with phenotypic manifestations of coagulation factors deficiency both in mouse and human	152
Table 8 - List of differentially expressed genes between different time points in the study.	178
Table 9 - Master regulator candidates from Motif analysis	182
Table 10 – Master regulator candidates from combining RNA-seq and Motif analysis	182
Table 11 – Transdifferentiation between somatic cell states	204

LIST OF SUPPLEMENTARY FIGURES

Supplementary figure 1 – Efficiency of reprogramming of different cell lines using episomal method.....	205
Supplementary figure 2 - FACS characterization of HLC day 4 using endodermal specific markers CXCR4 and SOX17.....	206
Supplementary figure 3 - RNA sequencing analysis.....	209
Supplementary figure 4 – Gene expression of cells during hepatic differentiation protocol.....	210
Supplementary figure 5 – Immunostaining of cells during hepatic differentiation protocol.....	207
Supplementary figure 6 – Endoderm differentiation in iPSC from colony format.....	205
Supplementary figure 7 – Liver of mice treated with CCL4	211
Supplementary figure 8 - Expression of surface markers assessed by FACS	212
Supplementary figure 9 – Teratomas from HLCs at different time points	213
Supplementary figure 10 - Coagulation factors gene expression data in BJ iPSC line hepatic differentiation.....	216
Supplementary figure 11 - Coagulation factors expression in mouse embryonic liver samples in GEO dataset	218
Supplementary figure 12 - Pictures of mouse embryos during dissection.....	218
Supplementary figure 13 – Schematic representation showing how a TFBS binding at a specific sequence pattern is matched and scored	221

LIST OF SUPPLEMENTARY TABLES

Supplementary table 1 - In-house Database of protein coding genes with annotations for tissue specificity and transcription factors. The DB is built in MySQL using Relational Tables to match IDs and downloadable as .csv and .xml for analysis in R.....219

ABBREVIATIONS

2D – Two dimensional
3D – Three dimensional
AAV – Adeno associated virus
AFP – Alpha feto protein
BAL – Bio-artificial liver
BME – β mercaptoethanol
CCl₄ – Carbon tetrachloride
CM – Conditioned medium
COAT – COllagen And Thrombin
CSC – Cancer stem cell
CLL – Chronic lymphocytic leukaemia
DE – Definitive endoderm
DNA – Deoxyribonucleic acid
E – Embryonic day
E8 – Essential 8
EB – Embryoid body
EBNA – Epstein-Bar nuclear antigen-1
ECC – Embryonal carcinoma cells
ECM – Extracellular matrix
EMT – Epithelium-mesenchymal transition
ESC – Embryonic stem cell
F – Factor
FACS – Fluorescence-activated cell sorting
FAH – Fumarylacetoacetate hydrolase
FDA – Food and drug administration
FPKM – Fragments per kilobase of transcript per million fragments mapped
GEO – Gene Expression Omnibus
GFP – Green fluorescence protein

GMP – Good manufacturing practise
h – Human
H&E – Hematoxylin and eosin stain
HDAC – Histone deacetylase
HFF – Human foreskin fibroblast
HLA – Human leukocyte antigen
HLC – Hepatocyte-like cell
HSC – Haemotopoietic stem cell
ICM – Inner cell mass
IP – Inter peritoneal
iPSC – Induced pluripotent stem cell
ITR – Inverted terminal repeat
IV – Intra venous
LIF – Leukemia Inhibitory Factor
LQTS – Long QT syndrome
MEF – Mouse embryonic fibroblasts
miRNA – Micro RNA
miR – Micro RNA
mRNA – Messenger RNA
ORA – Over-representation analysis
OSKM - Oct4, Sox2, Klf4 and c-Myc
OSNL - Oct4, Sox2, Nanog and Lin-28
PARS – Protease-activated receptors
PAS - Periodic acid-Schiff
PCR – Polymerase chain reaction
PB – PiggyBac
Pc – Post conception
PPI – Protein-protein interactions
PSC – Pluripotent stem cell
RA – Retinoic acid
rAAV – Recombinant adeno-associated virus
RDP – Rapid onset parkinsonism dystonia

RNA – Ribonucleic acid
RNA-seq – RNA-sequencing
RPE – Retinal pigment epithelial
RT-PCR – Reverse transcriptase PCR
SCNT – Somatic cell nuclear transfer
sh – Small hairpin
siRNA – Small interfering RNA
ss – Single-stranded
SV40LT – SV40 large T antigen
TERT – Telomerase reverse transcriptase
TF – Transcription factor
TFc – Tissue factor
TFBS – Transcription factor binding site
TGF- β 1 – Transforming growth factor β
VST – Variance stabilizing transformation
vWF – von Willebrand factor

Chapter 1 Literature review

1.1 Stem cells

Stem cells are cells that have the ability to self-renew and to differentiate into multiple lineages. This concept was first described in the 1960s, when James Till and Ernest McCulloch showed that the injection of bone marrow cells into irradiated mice led to multiple lineage hematopoietic colonies in the spleen. While not known at the time, the first transplant of hematopoietic cells was performed.

Humans and mammals in general, begin their lives with the fertilization of an egg by a sperm cell. The fertilization process results in a cell – the zygote - that has the potential of growing into a whole organism composed of very different cell types. The process by which cells go from an undifferentiated state to a specialized state is called differentiation. Waddington's model (Figure 1) helps explain the process by which a zygote differentiates into the multitude of cells that compose an adult organism, and has become key in the field of developmental biology.

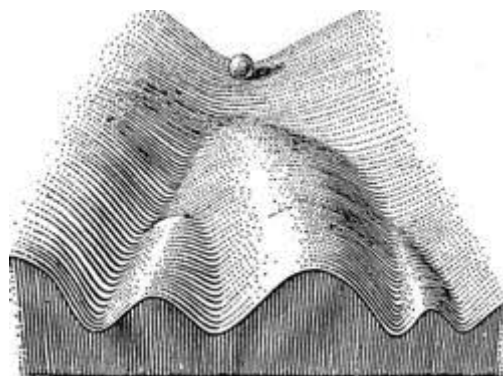


Figure 1 – Waddington's model on epigenetic landscape

The model shows a ball on top of a hill. As it rolls down, the ball can take several possible routes. This metaphor fits well with the specialization process cells undergo during development, as once a cell goes down a specific route it is hard for it to “go up the hill” or “across to another trough”. Limited options are possible after a route has been taken until the ball reaches the end of the hill, which depicts the different levels of differentiation ability cells have. And so, stem cells can be divided into 3 categories according to their differentiation potential: totipotent, pluripotent and multipotent. Upon fertilization and within the first cell divisions, cells are totipotent with the ability to form the developing organism as well as extra embryonic structures such as the placenta. Four-to-five days after fertilization, the zygote reaches the blastocyst stage where pluripotent cells located in the inner cell mass are encountered. These cells have the ability to differentiate into any of the 3 germ layers: endoderm, mesoderm and ectoderm, and therefore form every cell of an organism. Multipotent cells come into a more specialized area as they can only give rise to cells of a specific lineage/tissue type.

Due to their inherent characteristics, stem cells are considered to have invaluable potential as cellular resources for the treatment of several disorders. There is much interest surrounding pluripotent stem cells for these cells can be differentiated into any cell of the human body and we currently have the knowledge to develop such cell lines. Different pluripotent stem cells have been derived, among which embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) are included.

1.1.1 Embryonic Stem Cells

ESCs are the *in vitro* derivatives of the inner cell mass of the blastocyst. These cells preserve the pluripotency and self-renewal characteristics of the cells *in vivo* and allow the study of an organism’s early development. By harnessing the differentiating potential of pluripotent cells *in vitro*, the field of regenerative medicine observed the beginning of a new era.

In 1981, the first mouse ESC lines were established, and almost 2 decades later James Thomson was able to generate the first human ESC line¹. These cells have the ability to

differentiate into cells of the 3 germ layers and since their discovery, many protocols have been established to differentiate ESCs to specific cell types of all germ layers.

Despite all the potential these cells present for regenerative medicine, there are challenges that need to be overcome, primarily the ethical issue of destroying human embryos in order to retrieve ESCs from the inner cell mass. Furthermore, there is the challenge of obtaining suitable donor cells that are safe and effective. Additionally, tissue or cells to be used in patients need to overcome the problem of immune rejection.

1.1.2 History of reprogramming

Even before the concept of stem cell existed, reprogramming experiments by somatic cell nuclear transfer (SCNT) were being performed. In 1952, cells of late blastula were injected into an enucleated frog egg that resulted in the formation of a complete blastulae and, occasionally, more developed embryos². Even though none of the blastulae came to term, this was the first in a series of experiments where reprogramming by SCNT was successful^{3, 4}.

Decades later, a breakthrough came when the first successful reprogramming of mammalian cells was achieved with the generation of Dolly the sheep⁵. This was done by electrofusion of an adult sheep mammary epithelial cell with enucleated ewe egg followed by transplantation into a recipient animal. A great furor surrounded this scientific achievement as it was the first instance of mammalian cloning. Since then, other successful cloning of mammalian species have been achieved, including mice, rabbit, pigs and many others⁶⁻⁹.

Even though the process is incredibly inefficient, the SCNT experiments demonstrated that oocytes contain the factors necessary to reprogram differentiated cells. Considering Waddington's model, these experiments showed it was possible to revert cells "up the hill" back to a pluripotent state (Figure 2).

Developmental potential

Totipotent
Zygote

Pluripotent
ICM/ES cells, EG cells,
EC cells, mGS cells
iPS cells

Multipotent
Adult stem cells
(partially
reprogrammed cells?)

Unipotent
Differentiated cell
types

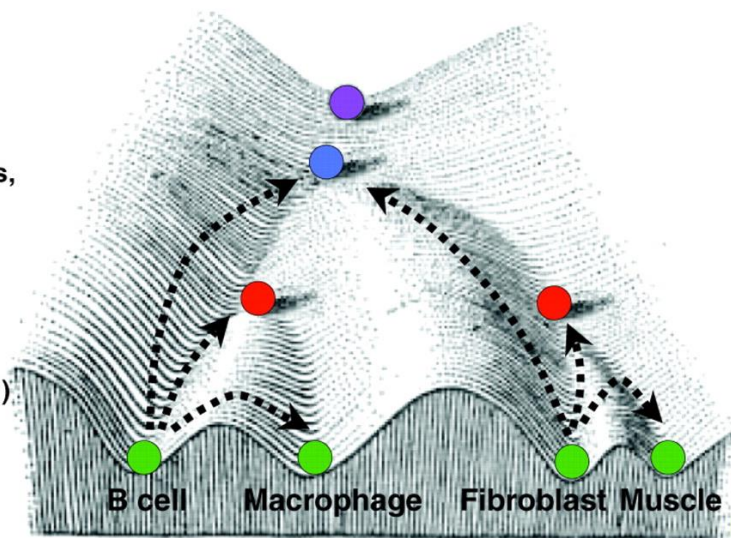


Figure 2 – Modification of Waddington's epigenetic landscape showing reprogramming possibilities of differentiated cells. Model showing cell populations at different developmental stages according to Waddington's model. Development potentials shown on the left with examples of cells at each stage. Figure from Hochedlinger and Plath¹⁰.

In 1976, Miller and Ruddle showed the first reprogramming by cell fusion in which fusion of diploid murine embryonal carcinoma cells (ECCs) and diploid thymocytes resulted in a hybrid cell where the phenotype of pluripotent stem cells (PSCs) dominated¹¹. The cancerous phenotype of ECCs led to the replacement of these cells by ESCs, which showed similar abilities in reprogramming by cell fusion¹². Even though the hybrid cells possess characteristics of PSCs, such as teratoma formation, they are tetraploid or near-tetraploid containing DNA from both cell types, which greatly impedes their use. Cell fusion experiments with ESCs showed that ESCs possess reprogramming capabilities similar to that of oocytes.

SCNT and fusion studies suggest that both oocytes and ESC possess factors that allow the reprogramming of somatic cells. This led to the hypothesis that cellular extracts from these cells can induce reprogramming of somatic cells.

In 2005, reprogramming with cellular extract was achieved, by which HEK 293T and NIH3T3 cells were permeabilized, incubated with ESC extract and resealed¹³. The cells

changed and were reprogrammed to a pluripotent state by which they present similar morphology to ESC, the pluripotency gene OCT4 was found upregulated and showed ability to differentiate into other lineages in presence of retinoic acid (RA)¹³.

The evolution of reprogramming allowed us to understand that there are factors present in PSC that allow the reprogramming of somatic cells to a pluripotent state. This knowledge was the basis for the development of iPSCs.

1.2 iPSC reprogramming

In 2006, the first iPSC were derived. These were obtained by the over-expression of four transcription factors: Oct4, Sox2, Klf4 and c-Myc (known as Yamanaka factors or OSKM) by retrovirus-mediated transduction of mouse fibroblasts¹⁴ (Figure 3). This was a revolutionizing study by which somatic cells were induced into a pluripotent state by the introduction of exogenous factors. Since then, the interest in this technology has boomed greatly due to its potential in regenerative medicine and disease modelling. The increased interest was accompanied by expansion on the reprogramming methods.

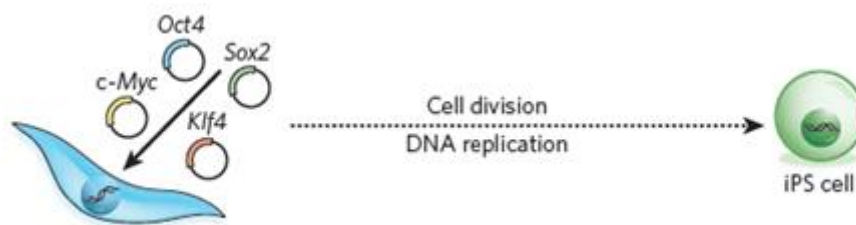


Figure 3- Reprogramming process. Simplified schematic of the reprogramming of somatic cell into iPSC state by introduction of the OSKM factors.

These cells have potential in several fields of research and despite the multitude of methods available we should be aware how different factors can influence the reprogramming process and the quality of the final population. The donor cell type, the reprogramming cocktail, the culture condition and the delivery systems are the main ways by which the reprogramming process can be affected.

1.2.1 Donor cell type

The donor cell type greatly influences the efficiency and kinetics of the reprogramming process. The most common cell type used are fibroblasts, but when we compare efficiency of

mouse embryonic fibroblast (MEFs) to human foreskin fibroblast (HFF) we observe an almost 2 fold increase in the time required for reprogramming. This difference in the kinetics of reprogramming might result from species differences in cells but most likely due to the different developmental stage of the cells: embryonic vs adult.

The effect of the differentiation status of cells on the reprogramming process is also shown by Eminli *et al.* where haematopoietic stem cells (HSCs) generate 300x more iPSC colonies than terminally differentiated B and T cells¹⁵. The differentiation status of cells relies on their epigenetic state. Epigenetic modifications allow differential expression in otherwise genetically identical cells in the human body, and thus provide a unique signature of a cell differentiation status. The more differentiated cells are, the more repressive marks they have, locking them into a specific status. The pluripotent state requires an open epigenetic state, and so during the reprogramming process the repressive marks need to be removed so that the iPSC state can be reached. This refers back to Waddington's model (Figure 1 and Figure 2) by which the amount of epigenetic markers in the cells correlates to their differentiated state and so, as cells "go down the hill", they acquire marks locking them into specific lineages. This explains why less differentiated cells have faster kinetics of reprogramming.

The donor cell type can also affect the composition of the reprogramming cocktail as certain donor populations are known to endogenously express certain reprogramming factors which obviates their expression in *trans*. This is exemplified in melanocytes: as these express high levels of endogenous SOX2, reprogramming can be done with only 3 factors¹⁶. Another example that shows the extent of how the cell type can affect the reprogramming cocktail was reported by Scholer *et al.* when it was demonstrated that the sole introduction of OCT4 in fetal neural stem cells was enough to achieve reprogramming¹⁷.

As shown, several cell types besides fibroblasts have been used for the reprogramming process, including blood cells. Reprogramming from blood cells is highly appealing as cells can be collected in a non-invasive procedure for patients, already routinely used in the clinic, which is advantageous when compared to skin biopsies required to obtain fibroblasts. Unfortunately, the efficiency observed in differentiated T and B cells is reduced when compared to HSCs. Availability of cell type also affects their use for reprogramming, as in

the case with HSCs, and so, until higher efficiencies can be obtained from differentiated blood cells, fibroblasts might remain the preferred cell type for reprogramming.

Different features affect the suitability of each cell type for the reprogramming process, with each having its own advantages and limitations. It can thus be suggested that the ideal cell type for reprogramming should be (1) easily accessible, (2) with minimal risk in obtaining the cells from patients, (3) available in large quantities with relative high efficiency and (4) speed of reprogramming should be taken into consideration.

1.2.2 The reprogramming cocktail

The initial cocktail for reprogramming comprised the OSKM factors but other combinations have been used, often using factors expressed early during development and involved in the maintenance of pluripotency. Such factors include Nanog, UTF1 or SALL4 that, when added to the core reprogramming factors, increase efficiency and reduce the reprogramming time¹⁸.

Other factors such as MYC and KLF, have shown to affect cell proliferation, both directly or indirectly. Increased cell proliferation expedites the appearance of colonies, increasing the kinetics of the process. Telomerase reverse transcriptase (TERT) and the SV40 large T antigen (SV40LT) are proteins with known effects on proliferation, and indeed their positive effect on the reprogramming process has been shown¹⁹. Chemical compounds can also be used to aid in cell proliferation, and consequently in the reprogramming process. The use of chemical compounds extends beyond their effect on cell proliferation. Different compounds can affect a range of different cellular mechanisms, with modifications at the epigenetic level showing great improvements in the reprogramming process in various cell types. The epigenetic state of the donor cells is a known limitation, mainly due to the chromatin remodeling step²⁰. Chromatin remodeling can be facilitated by altering DNA methylation status or by modifying chromatin, which can be achieved by using histone deacetylase (HDAC) inhibitors (such as hydroxamic acid and valproic acid)²⁰. Some compounds can even have additive effects on the reprogramming, such as the case with Vitamin C that has shown

to improve the process both by inducing DNA demethylation²¹ and by alleviating cell senescence²².

The reprogramming cocktail can be affected by the donor cells used as each might have different requirements for reprogramming, as it has been described that some cell types can endogenously express some of the factors which limits their requirements in the reprogramming cocktail.

MicroRNAs (miRNAs) have been shown to support iPSC induction, both in combination with other factors and by themselves²³. These miRNAs have been identified by being preferentially expressed in ESCs and are thought to help maintain the pluripotent state^{24, 25}. This avenue is highly appealing as it allows to regulate gene expression without the need of protein elements, reducing the processing time and effort of cells.

1.2.3 Culture conditions

The environment cells are cultured also affects the reprogramming process. Generally, iPSC are kept in identical conditions to ESC, which are permissive to the pluripotent state.

One of the most influential variants in terms of culture condition is the use of a supportive cell layer. These cells are known to secrete factors required for PSCs survival, proliferation and pluripotency maintenance and often MEFs and neonatal human dermal fibroblasts are used. The risk of contamination from the feeder layer led to development of supportive systems without the use of these cells. A range of feeder-free systems have been developed, be it protein extracts, recombinant proteins or even synthetic matrices²⁶⁻³². These have varying degrees of efficiency in maintaining stem cell culture and supporting the reprogramming process. Some only work in combination with specific media, which shows the importance of the combined effects of culture conditions for proper iPSC regulation.

Medium composition can also affect the reprogramming and while in early times it relied mostly on the factors secreted by the feeder layer, it has come a long way since then. Better

understanding of the signals required to maintain pluripotency allowed to develop more defined media that contain all the necessary elements to maintain the correct phenotype of cells.

Changing the conditions so that these resemble the natural niche of cells has been suggested as a way to improve the culture conditions. Indeed, when cells were cultured in hypoxic conditions, reprogramming efficiencies of both mouse and human iPSC showed an increase³³. This was based on the knowledge that stem cell niches often possess lower O₂ % than the atmospheric 21% O₂.

Others have suggested the effect of stress conditions in the reprogramming of somatic cells. This process has been observed in plants, by which the exposure to drastic environmental changes can convert cells from a mature state to an immature state³⁴. Although a paper came out showing how low-pH conditions were amenable to acquisition of pluripotency³⁵, it was later retracted³⁶ and so no evidence exists that environmental effects can solely reprogram cells into a pluripotent state. Even so, it has been shown that environmental conditions aid the reprogramming process, as in the case of hypoxic conditions. It is known that environment conditions can regulate gene expression and so it is possible that other environmental cues can help improve the reprogramming process.

1.2.4 Delivery systems

No other field observed such variety as the delivery systems. A method to introduce the reprogramming cocktail in the cells is needed, and while the initial study relied on retrovirus many other ways have been used to do so. We can subdivide the wide range of methods available into two important categories: integrative and integration-free.

Retrovirus and lentivirus are two common viral vectors used to generate iPSC and both are integrative systems. As the name suggests, the transgenes are incorporated in the genome of the host cell, and in general have higher reprogramming efficiencies as the integration in the genome allows consistent expression of the reprogramming factors. This expression, even if

advantageous for the reprogramming process, creates a problem for the resulting iPSC in terms of safety. Genomic integration of exogenous DNA can affect the stability of the genome and depending on the site of integration it can affect tumour suppressor genes or oncogenes – issues already seen in a gene therapy study using retrovirus³⁷. Furthermore, constitutively expression of the reprogramming factors after the reprogramming process also poses risks as some of the factors are potent oncogenes (such as c-Myc).

While in the case of retrovirus the transgenes are generally silenced³⁸, the risk of re-activation is still present. In the case of lentivirus, silencing of transgenes is less effective³⁹ which led to adaptation of drug-inducible systems to allow controlled expression of the factors^{40, 41}. This type of system improves the silencing efficiency but again the risk of re-activation is still present. The advantage of lentivirus compared to retrovirus is the fact that it allows transducing non-dividing cells and dividing cells with similar efficiencies³⁹.

To avoid the use of viral vectors, DNA transfections were used and showed a significant decrease of efficiency as lower numbers of cells underwent complete reprogramming. The design of polycistronic vectors allowed a slight increase in reprogramming efficiency⁴². The lower efficiency coupled with a still integrative method does not bring much advantages compared to viral methods. In an attempt to remove the integrated DNA from the host genome, vectors with loxP sites flanking the reprogramming cassette were used and shown to indeed induce pluripotency of transfected cells. Upon transient expression of Cre recombinase in these cells, the loxP flanked sites are deleted from the genome, which allows to remove the transgenes thus improving the cells overall safety by reducing oncogenic potential.

Even if the removal of transgenes can be achieved, the use of the loxP method still leaves genomic scars behind that could still disrupt expression of tumour suppression genes or oncogenes. The only integrative method shown to be able to remove the inserted DNA with no scars behind are vectors based on the PiggyBac (PB) transposon⁴³. While it allows complete removal of transgenes, DNA alterations have been reported and therefore sequence verification is needed¹⁹.

The development of integration-free reprogramming methods removed these issues and made the final iPSC population safer overall. Several methods exist and their general disadvantage compared to integrative methods is lower reprogramming efficiencies (Figure 4).

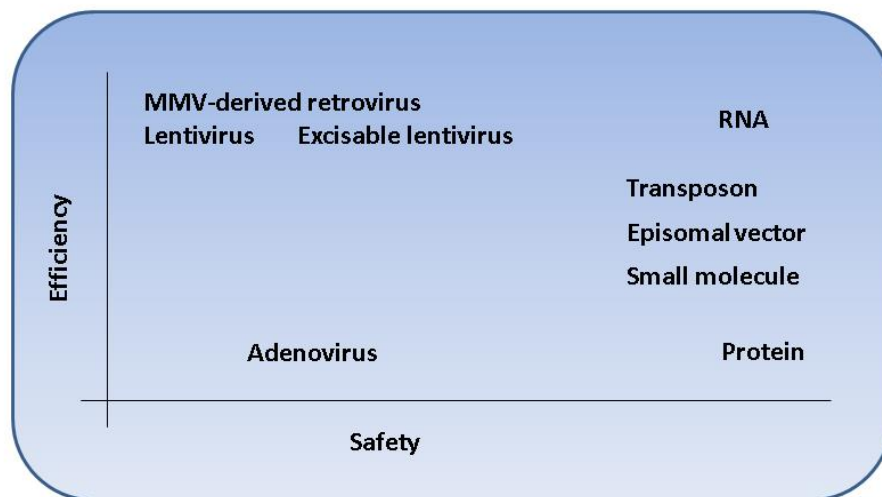


Figure 4 – Graph showing the relation between efficiency and safety of different reprogramming methods. Adapted from González *et al.*⁴⁴

One of the first attempts at integration-free methods, and with some of the lowest efficiencies, was using integration defective viral delivery by using replication-defective adenoviral vectors⁴⁵. Despite the ability to create iPSCs, it was only successful when hepatocytes were the donor cell type, requiring integrative expression of factors for other donor cells⁴⁵. This shows us once more how the donor cell type can influence the reprogramming process, not only by itself but in combination with all other factors that affect the reprogramming process.

Another viral vector used in integration-free manner is the Sendai virus. These viruses belong to the RNA virus family that replicate in the cells' cytoplasm, which means no genomic integration occurs⁴⁶. Using F-deficient sendai virus reprogramming of human fibroblasts and T cells has been successfully achieved^{47, 48}. Despite the safety measures to prevent viral replication, it was difficult to eliminate viral vectors from the host cells and, even at higher

passages, clones with transgene expression are detected⁴⁷. And so, even if sendai virus are integration-free, the late detection of transgene expression creates a hurdle to their use.

To avoid the use of viral particles, researchers turned their focus to other integration-free options such as episomal plasmids. These allow the transient expression of transgenes in the host cells. Expression allows reprogramming to occur and is subsequently lost, allowing the removal of any exogenous material in the cells. While non-replicating episomal plasmids showed success in reprogramming MEFs, they proved unable at reprogramming human cells⁴⁹. The inefficiency at reprogramming human lines is probably due to the short window the transgenes are expressed, which coupled with low transfection efficiencies and dilution of factors as cells proliferate makes it even harder for cells to undergo changes that lead to pluripotency. Even when successful, non-replicating plasmids showed cases where integration into the genome occurs^{49, 50}.

To try and circumvent this issue, Yu *et al.* used oriP/Epstein-Bar nuclear antigen-1 based episomal vectors (oriP/EBNA1)⁵¹. These vectors can be kept under selection conditions and, in the absence of drug selection, are lost at a 5% rate per cell division. The longer periods allows for the reprogramming process to occur. The low efficiencies led Okita *et al.* to optimize the process by introducing elements that increase reprogramming efficiency: p53 small hairpin and L-Myc instead of c-Myc⁵².

To try and improve the low transfection efficiencies that are observed with episomal vectors, minicircle vectors were developed. These are super-coiled DNA molecules that lack a bacterial origin of replication, which in turn gives them longer transgene expression times as it helps evade exogenous DNA-silencing mechanisms^{53, 54}. The higher transfection efficiencies allows the mini circles to compensate for serial dilution and for reprogramming to occur⁵⁵.

Replacing genetic-based techniques by other methods represents an important step in the search for safer methods. The use of RNA or proteins are viable alternatives as, not only they remove the risk of genomic integration, but allow to reduce the cellular processes for

transgene expression (transcription and translation), reducing the time necessary for transgene action.

Warren *et al.* were able to achieve high efficiencies of reprogramming with faster kinetics when using synthetic mRNAs⁵⁶. Their success depended on two important features: (1) modifications of the mRNAs in order for these to avoid the cells defense mechanisms against RNA virus and (2) daily transfections required to allow sustained expression of the factors as mRNA molecules have a fast turn over time.

The use of proteins is more challenging than RNA as the mechanisms for their delivery are not as easy, relying on peptides fused to the reprogramming factors that mediate their transfections^{57, 58}. The low efficiency and kinetics coupled with the difficulty in protein purification and high amounts required make it highly unlikely for routine use, but as proof of concept, it has been shown to work.

Overall, the rapid progress of the iPSC field was accompanied with improvements in iPSC derivation methodologies, and currently a wide range of options is available. Different factors can affect the reprogramming process and even though we presented them as separate fields, the reality is that these are not independent and should be taken together when determining the best method of reprogramming. The ideal reprogramming method should comply with safety and efficiency requirements, but ultimately the best reprogramming method may vary depending on the specific application of the iPSCs

1.3 Applications of iPSC

One of the main interests surrounding iPSC is the advantages it can bring to the field of cell transplantation. Various cell types can be obtained from PSCs and their unlimited proliferation ability allows to obtain any number of cells required to repair tissues damaged by disease or injury. The autologous nature of iPSC removes the risks of rejection and infection making them ideal for cellular therapies. Furthermore, it allows to overcome the ethical issues associated with the destruction of embryos for the acquisition of ESCs.

The therapeutic effects of ESC have been extensively described in animal models of several disorders, including spinal cord injury⁵⁹, retinal disease⁶⁰ and Parkinson's disease⁶¹. A proof-of-concept study with a humanized sickle cell anemia mouse model showed that iPSC could be used to treat single gene defects⁶². In this study, Hanna *et al.* corrected the sickle haemoglobin allele in iPSC by homologous recombination and upon differentiation into haematopoietic progenitors and transplantation into mice showed successful rescue of the disease phenotype⁶². Other studies have shown the feasibility of iPSC to treat disease or injured phenotypes⁶³⁻⁶⁷, and even though their clinical use is still at its infancy, two encouraging clinical trials using ESC-derived cells have been approved by the FDA (www.clinicaltrials.gov) and are paving the way for PSCs clinical use. The first one was for the treatment of spinal cord injury using ESC-derived cells while the second was focused on macular degeneration of the eye using retinal pigment epithelial (RPE) cells from ESCs. These studies are important milestones in the transition of PSCs therapies to the clinic, especially in safety matter as in the case of macular degeneration where no hyper proliferation, abnormal growth or immune rejection was observed in the patients⁶⁸.

Even if cellular therapies using iPSC are still in a preliminary phase, iPSC have shown useful in other areas. The ability to retain the genomic information from a specific patient allows the opportunity to study "disease in a dish".

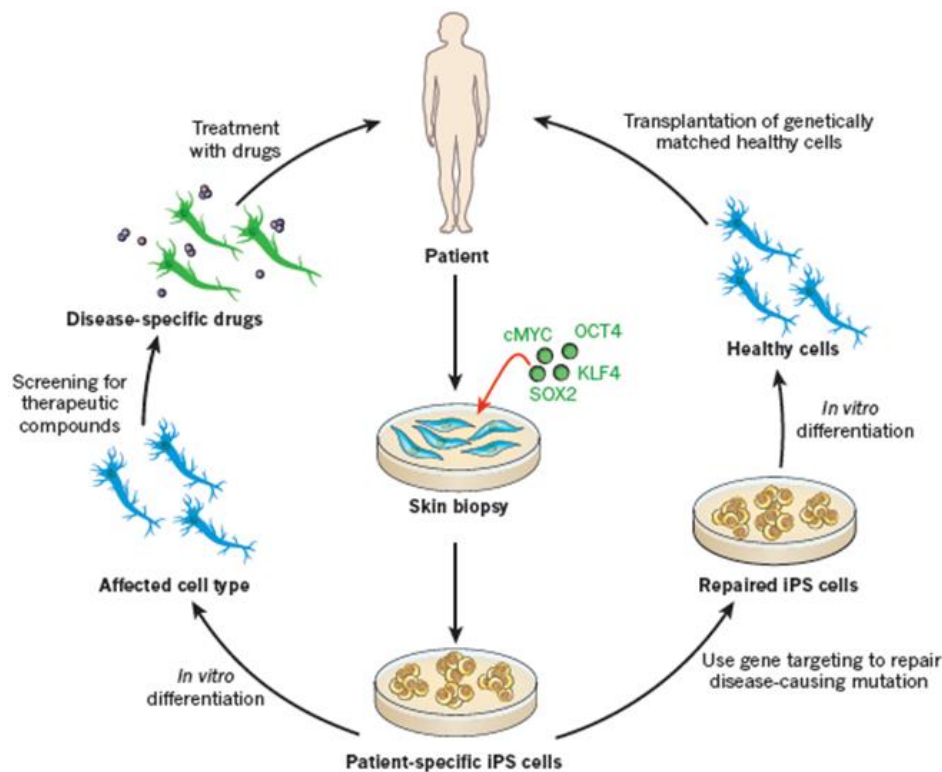


Figure 5 – Applications of iPSC. Possible applications of iPSC technology. Figure from Robinton and Daley⁶⁹.

Disease modelling gives the opportunity to study the development and functional implications of human diseases (Figure 5). The first model to study human pathogenesis was for the study of spinal muscular atrophy, where the patient derived iPSC showed the same phenotypical hallmarks of the disease when differentiated into motor neurons⁷⁰. The study of human neurological disorders benefited greatly with the development of iPSC technology as the complexity of the neuronal system and the difficulty in culturing neurons *in vitro* made it hard to study the development and function of human neurons⁷¹. iPSC have allowed not only to replicate the disease phenotype *in vitro*⁷⁰ but even aided in elucidating the molecular mechanisms underlying disease⁷²⁻⁷⁴. Neurological disorders are not the only ones that have been modelled using iPSC. To date, a variety of disorders affecting other tissues or organs have also been used, including immunological disorders and cancer⁷⁵⁻⁷⁷.

The *in vivo* niche where cells exist are very complex and determinant for their phenotype. *In vitro* culture systems have greatly improved in their ability to replicate the *in vivo* situation

but they are still short of a perfect system. For disease modelling, stress factors might be necessary to mimic the pressures cells are under in the disease state. Furthermore, improvements on the quality of cell differentiation are needed as seldom these give rise to homogeneous population of cells, with the most common issue being the correct synchronization of the developmental stage of cells in culture⁷⁸. Despite the possibility for improvements the usefulness of iPSC in disease modelling cannot be challenged.

By helping elucidate disease mechanisms, iPSC also open the door for toxicology screens. By comparing normal development with diseased state, the possibility of identifying new targets creates opportunities for new therapeutical agents⁷⁴. Parallel to toxicology screens, drug metabolism studies represent an important feature of iPSC. These take advantage of the ability to use specific genetic backgrounds as a starting point for hepatic or cardiac differentiation.

The majority of drugs fail to reach the market due to cardiotoxicity or hepatotoxicity. The current systems are limited in their ability to predict unforeseen effects (such as absorption, distribution, metabolism and even toxicity) which incurs high cost in the drug-screening process and consequently in the pharmaceutical industry. iPSC technology has proved innovating in modelling human physiology *in vitro* as it allows to use a large genetic pool and use both healthy and disease phenotypes as models for drug screening. In the liver context, individuals with different cytochrome p450 enzymes can show different metabolic profiles of drugs, which makes the use of different genetic backgrounds extremely valuable when assessing the effect and safety of drugs. The same concept can be used in the context of the heart, as certain drugs can affect the cardiac action potential and cause long QT syndrome (LQTS) which can result in lethal arrhythmias. The ability to differentiate iPSC into beating cardiomyocytes allows to test, in an *in vitro* setting, the sensitivity to drugs and their effect in the action potential of cells.

The use of human cellular products for drug testing is not new but the unlimited source of cells from iPSC allows overcoming the limiting problems of cellular supply.

This technology greatly surpasses the use of animal models as it reproduces human physiology in a way animals (even genetically modified) cannot. An example is shown in the case of different pathophysiological features seen in the hearts of transgenic mice versus human heart where variations in ion channel types and distribution results in faster heart rates and shorter action potential duration in the mouse⁷⁹. Despite the limitations, animal models continue to be important for understanding disease mechanisms and drug screening. In the future, better iPSC-derived models should be chosen over its animal counterpart, as often these do not truly mimic what happens in humans.

In addition, iPSCs are also emerging as helpful tools in the field of developmental biology. The differentiation of iPSC into somatic cells allows to replicate *in vitro* the natural processes. Even if the differentiation protocols are limited comparing to *in vivo* development, they allow us to tap into periods of development that otherwise cannot be studied.

Overall, iPSC technology has proven itself useful for research purposes with optimistic data regarding their use in cellular therapies. Despite the promises of this technology, there are challenges that still need to be addresses and considerations to be taken into account if this technology is to be used therapeutically.

1.4 Questions and challenges of the iPSC field

Since the first iPSCs were derived in 2006, the field has observed a progress at a breathtaking pace. With the development of the field the unanswered questions and challenges associated with the cells also increased. Some of the main issues that surround the iPSC field are presented here.

1.4.1 ESCs Vs iPSC

The iPSC field piggybacked on the ESC field, but the question on how equivalent these cells truly are still remains. Indeed, the development of the iPSC field relied on factors shown to maintain pluripotency in ESCs. It is then expected that iPSC and ESC should be equivalent. As anticipated, many if not all of the features of ESC are observed in iPSC. These are morphologically identical and show the same developmental potential as ESCs⁸⁰ with the ability to generate cells from the three different germ layers.

Other molecular assays, which compare gene expression and epigenetic characteristics allow a more quantitative comparison between the two. While initial reports showed that profile of global expression, histone modifications, X-chromosome inactivation and DNA methylation profiles of iPSC were similar, if not indistinguishable from ESC⁸¹, later studies contradicted these findings. These showed iPSC as having its own gene expression signature, different from ESC⁸²⁻⁸⁴. Differences in DNA methylation between the two type of cells were also identified^{85, 86}, with epigenetic memory of donor cell also reported⁸⁷⁻⁸⁹.

However, the analysis of a large collection of gene expression and histone modification data shows that the differences seen in the previous studies might represent “experimental noise”, by which there is an overlap in the variations between ESC and iPSC clones, thus clustering iPSC and ESC together^{90, 91} (Figure 6).

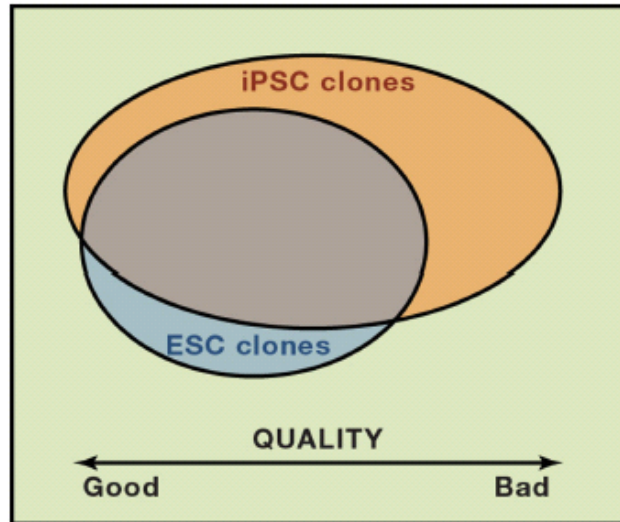


Figure 6 - Overlapping quality between iPSC and ESC. Variations observed between iPSC and ESC regarding a range of properties (including gene expression, DNA methylation and differentiation propensity). While most lines are grouped together there are some iPSC that differ from ESC. Figure from Yamanaka⁹².

While these studies cluster iPSC and ESC together there are questions regarding epigenetic memory in iPSC, especially observed in early passage lines⁹³. Soldner *et al.* proposed that the changes in gene expression might just result from residual transgene induction that are specific to cell of origin⁹⁴. This shows that there are factors external to iPSC but relevant for its induction that can affect how much they resemble ESC. Some of the parameters that complicate the comparison include presence of incomplete silencing of transgenes, different reprogramming cocktails, genetic background of cells, natural heterogeneity among ESC lines, and incomplete reprogramming to pluripotent state⁹⁵.

How important are the differences between iPSC and ESC for the differentiation of functional cells is not clear as tolerance towards some epigenetic and gene expression aberrations has been shown. The level that is tolerated is another question that might depend on how closely related the lineages of the target and donor cell type are.

The use of ESC as the standard control for iPSC quality makes sense as iPSC were developed from knowledge obtained in ESC. Nevertheless ESCs are seen as *in vitro* cells of the inner cell mass (ICM), while they - similarly to iPSCs - are a man-made cell type. Changes

between ESC and cells of the ICM have also been reported, both at the methylation level⁹⁶, gene expression level⁹⁷ and even regarding the length of telomeres⁹⁸. And so, both the differences and similarities of iPSC with ESC can have implications in their quality that escape our knowledge.

1.4.2 Safety concerns with iPSC

Concerns regarding the quality and safety of these cells remain the main focus in the iPSC field.

As discussed previously, much has been done to improve the safety of iPSC when it comes to the reprogramming method. Besides the safety of the method one must consider efficiency and consistency of manufacture in order to ensure reproducibility of the process. This is an important factor in the manufacturing process as regulatory concerns can hold the clinic use of these cells.

In addition to the aforementioned, there are other features of iPSC that need to be considered in terms of their safety. We previously saw that iPSC and ESC show differences at the genetic and epigenetic level, with implications still unclear. Even if the differences between these two cell types are not significant, the implications regarding their forward differentiation is crucial information for their use.

Genetic stability is an important feature to any cell, as instability is one of the hallmarks of cancer⁹⁹. While genetic instability of human ESC and iPSC in culture has been shown^{100, 101}, evidence is starting to emerge suggesting it results from *in vitro* culture rather than the pluripotent state, as instability was also detected in adult stem cells in culture^{102, 103}. Karyotype profiling might not be sufficient, and more precise genetic analysis should be done¹⁰⁴. As the risk of genetic instability increases with time in culture^{105, 106}, low number of passages might be required to decrease the risk of genetic alterations, however low passage numbers might implicate presence of transgene in the clones. As a minimum time in culture is required for the pluripotent state to be reached, optimization of the time in culture is

required. Other genetic differences among iPSC have been shown including somatic mutations¹⁰⁷, copy number variations (CNVs)¹⁰⁸ and immunogenicity¹⁰⁹. Nevertheless, some studies have shown these variations were there prior to the reprogramming process^{110, 111}. An acceptable threshold of genetic changes should be determined as some of these can be natural occurring differences that do not affect quality of cells (such as CNVs).

It is also crucial to assess the heterogeneity of a culture as there is tumorigenic or immunogenic risks associated with the engraftment of undifferentiated or incorrectly differentiated cells. Tumours following stem cell use have been reported¹¹² demonstrating the tumorigenic risk associated with PSCs. It is not easy to quantify the tumorigenic risk of a cell population and while the pluripotency inherent to PSCs is a risk, it is its cellular derivatives that have interest for therapeutic use. The tumorigenic risk is still present either from instability correlated with time in culture or from quality of differentiation protocols. This risk can be decreased with increased purity of population (either with improved protocols or by purification of the final population) as well as by monitoring for contamination. The study of biodistribution of iPSCs and its derivatives must take into consideration a range of factors such as the route of administration, methods for cellular detection and their sensitivity, number of cells, as well as the species used as models. Different methods for *in vivo* tracking are available but “sensitivity vs specificity” has made it hard to conclusively determine the biodistribution of cells *in vivo*. Animal studies for assessment of human PSC derivatives generally rely on use of immuno-suppressed or immuno-compromised animals, but masking the effect of the immune system could also play a role in the biodistribution of cells, further complicating the significance of any findings.

Although the use of iPSC technology should remove the risk of immunogenicity, an initial study by Zhao *et al.* showed a T-cell dependent immune response when mouse iPSC were used in syngeneic recipients¹⁰⁹. Cellular therapy will rely on PSC derivatives rather than on PSC in their native state and so, even if the study by Zhao *et al.* begs caution in terms of an immune response, similar studies with the use of differentiated cells were needed. Such studies have shown contradicting data, some showing similar results to Zhao where iPSC

derivatives produced an immune response¹⁰⁹ while others claimed “negligible” or “lack of” immunogenicity¹¹³. These studies suggest that not all tissues derived from iPSC elicit an immune response, and that these might be dependent on the quality of the starting iPSC population and/or final population. In general, iPSC still present a better alternative to ESC when considering immunological issues as shown by a study where differences in the immunogenicity of endoderm cells derived from iPSC and ESC were observed¹¹⁴. Quality of the iPSC population and maturation status of the cells are not the only factor that can affect immunogenicity, with other factors such as site of administration, number of doses, the immunological basis of the disease and even the age of the recipient being able to affect the immune response to treatment¹¹⁵.

1.4.3 Quality of differentiation protocol

Quality and maturation state of the iPSC derivatives are important in terms of safety (tumorigenic risk) but go beyond the use of iPSC for the clinic. The iPSC field has applications in disease modelling, drug toxicity and even developmental biology. For any of these, the results obtained are as good as the quality of the differentiation protocol and the final population of cells.

In vitro differentiation protocols aim to mimic the developmental stages for each specific lineage, but fall short of perfect systems. Most protocols result in a heterogeneous population of cells that present a foetal phenotype. A great deal of research has been done to improve those, either by use of small molecules¹¹⁶ or better culture systems (such as in the cytoarchitecture of cells) during the differentiation protocol.

Small molecules have the added advantage of being cheaper and easier to produce than recombinant protein factors and show higher ease of scalability¹¹⁷. Another feature specific to small molecules is the ability to inhibit rather than activate signalling pathways. This has been shown in neural specification¹¹⁸, with highly synchronized and efficient induction. Similar approaches were seen for other lineages¹¹⁹.

Three-dimensional (3D) structures allow cells a more physiological cytoarchitecture, a crucial feature for several cell types. During development, cells interact not only with each other but with extracellular matrix (ECM), interactions important for multiple functions such as cell growth, differentiation, cell maintenance and even tissue morphogenesis¹²⁰. This lead to increase interest in developing systems that mimic the natural niche of cells during embryogenesis¹²¹⁻¹²³, as these can help modulate the behavior of cells. A great deal of work has been done to improve the cytoarchitecture of cells but limitations still exist and need to be considered. Collaboration between the fields of developmental biology, stem cell and biomaterials will be key for development of better and more suitable culture systems.

Another consideration to take into account is the choice of cell source. We have seen that epigenetic memory can exist in iPSC, and while this represents a downside in the search for a pluripotent cell, it might be beneficial when considering differentiation of cells into a specific lineage. The donor cell type can influence the differentiation potential of iPSC and so one might consider using iPSC from related lineages when differentiating them. This is a controversial idea as truly pluripotent cells should not have differentiation bias toward any lineage, and thus can put into question the quality of the iPSC population.

The knowledge on iPSC is continuously increasing and becoming ever more complex. The more is understood, the more questions arise that need to be answered. Despite this, the interest in this technology has not faded, with growing hopes for its prospective clinical use.

Chapter 2 Materials and methods

2.1 Cell culture

Tissue culture was performed in level 2 safety hoods under sterile conditions. All media were sterile filtered and supplemented with 1mg/ml Penicillin/Streptomycin (P/S, Sigma). Solutions were pre-warmed at room temperature prior to use.

Table 1 - List of fibroblast cell lines used for reprogramming

Cell line	Description
RDP1	Skin fibroblast cell line from a 49 year old patient with rapid onset Parkinsonism dystonia (RPD)
RDP2	Skin fibroblast cell line from a 31 year old patient with rapid onset Parkinsonism dystonia (RPD)
IMR90	Fibroblasts cell line derived from the lungs of a 16-week female foetus
BJ	Fibroblasts cell line derived from neonatal foreskin

RDP1 and RPD2 cell lines were obtained from skin fibroblast from a 49 and 31 year old patients, respectively, with rapid onset Parkinsonism dystonia (RPD). We used these cells for the induction of iPS cells with a maximum of thirteen passages in culture. IMR90 fibroblasts from the lungs of a 16-week female foetus and BJ fibroblasts from neonatal foreskin were also used for the induction of iPS cells. These human fibroblast lines and Platinum A amphotrophic packaging (PLAT A) cells were maintained in MEF medium: Dulbecco's modified eagle medium with L-glutamine (DMEM, Gibco Life Technologies) containing 10% heat-inactivated foetal bovine serum (HI-FBS, Gibco Life Technologies) and 1mM Non Essential Amino acids (NEA, Life Technologies) in a 5% CO₂ humidified atmosphere at 37°C . Human HEK 293 cells, HUH7 and HepG2 cells were cultured in D10 medium:

DMEM supplemented with 10% HI-FBS, in a 5% CO₂ humidified atmosphere at 37°C. For passaging, medium was discarded and cells washed twice with phosphate buffered saline without calcium and magnesium (PBS, Life Technologies). TryPLE (Life Technologies) was added and cells were incubated at 37°C for approximately 5 min. Cells were resuspended with medium and centrifuged for 5min at 400x g. The supernatant was discarded and cells transferred into a new flask with fresh medium. Cells were passaged every 2-3 days using a split ratio of 1:3.

iPSCs were generated and maintained in irradiated newborn human foreskin foetal fibroblasts feeders with mTeSR1 (STEMCELL technologies) or human ES medium: KO-DMEM (Life Technologies) containing 20% Knockout serum replacement (KSR, GIBCO), 1mM GlutaMax (Life Technologies), 1mM NEA and 0.1mM β -mercaptoethanol (Invitrogen). hES medium was supplemented on the day of use with 4 ng/ml recombinant human basic fibroblast growth factor (bFGF, Life Technologies). Medium was refreshed every day. For passaging, human iPS cells were washed once with PBS and then incubated with 1mg/ml Dispase (STEMCELL technologies) at 37°C for ~7min. When edge of colonies started to curl, dispase was removed and cells washed with PBS. Appropriate volume of mTeSR1 or hES was added and cells were scraped off the plate using a 5ml short glass pipette and transferred to a new dish with Nuff feeder cells. The split ratio was routinely 1:3.

For feeder-free culture and induction of iPS cells, cells were maintained in one of the following combinations: matrigel (Corning)/geltrex (Life technologies) and mTeSR1 or recombinant vitronectin (Life Technologies/STEMCELL technologies) and essential 8 (E8) medium (STEMCELL technologies). Plates were coated with matrigel/geltrex and kept at 4°C overnight. Before use, plates were warmed to room temperature. For dispase passaging, cells were washed once with PBS and then incubated with 1mg/ml dispase. When the edges of the colonies began to curl, dispase was aspirated and cells washed with PBS. Fresh medium was added and cells scraped off the plate using a 5ml short glass pipette (Corning). Disaggregated colonies were transferred immediately to matrigel/geltrex coated plates. The medium was changed daily. For EDTA passaging, cells were washed once with PBS and then incubated with gentle dissociation reagent (STEMCELL technologies). After 3-5 mins, EDTA was aspirated and fresh medium added. Using a 1000 μ l pipette tip, cells were washed

off the plate and gentle pipetting was performed to break up the colonies. Disaggregated colonies were transferred immediately to matrigel/geltrex coated plates. Medium was changed daily. The split ratio was routinely 1:6 (range between 1:3 and 1:10) for dispase method and 1:10 (range between 1:6 and 1:20) for gentle dissociation method. Similar process for E8 + vitronectin conditions.

2.1.1 Cryopreservation of hPSCs

For long-term storage cells were cryopreserved in liquid nitrogen using bambanker freezing medium (Anachem). Cells were collected in small clumps as described above and spin down for 5 minutes at 200x g. The cell pellet was resuspended in 500µl of bambanker freezing medium and transferred into cryovials (Greiner Bio-One). Cells were immediately transferred to -80°C in a Mr. Frosty (Sigma Aldrich) to achieve a rate of cooling of -1°C/minute. After 24h in a Mr. Frosty vials were transferred into liquid nitrogen where they were stored until use.

2.1.2 Thawing hpSC

Cryopreserved stem cells were thawed in a 37°C water bath until a small fraction was still frozen. 6ml of prewarmed mTESR1 medium was added drop-by-drop and cells were spin down for 5 mins at 200x g. Medium was aspirated, cells resuspended in fresh mTESR1 and transferred to matrigel/geltrex coated wells.

2.2 Cloning

miR302 cluster flanked with SgrAI cut sites was amplified from mouse genomic DNA by PCR and cloned into TOPO (life technologies K4500-02). miR302 fragment was cut out of TOPO-miR302 using SgrAI (NEB) and introduced into the SgrAI site of the rAAV vector.

miR302 cluster flanked with BamHI cut sites was amplified by PCR and cloned into TOPO (life technologies K4500-02). Shp53 fragment was removed from pCXLE shp53 hOCT4 plasmid by restriction enzyme digest using BamHI (NEB) resulting in pCXLE hOCT4. miR302 fragment was cut out of TOPO-miR302 using BamHI (NEB) and was introduced into the BamHI site of pCXLE hOCT4 vector.

All cloned plasmids were checked by restriction enzyme digestion and sequencing. All restriction enzyme digests were done according to manufacturers' instructions.

2.3 Reprogramming

2.3.1 Using retrovirus

PLAT A were plated at 2×10^6 cells per 100 mm dish and incubated overnight. Next day, the cells were transfected with pMXs vectors with Fugene 6 transfection reagent (promega E2311). Twenty four hours after transfection, the medium was collected as the first virus-containing supernatant and replaced with a new medium, which was collected after twenty-four hours as the second virus-containing supernatant. Human fibroblasts were seeded 10^5 cells per well of a 6 well plate in MEF medium 1 day before transduction. The virus-containing supernatants were filtered through a 0.45 mm pore-size filter (Merck Millipore). Equal amounts of supernatants containing each of the four retroviruses were transferred to the fibroblast dish, and incubated overnight. Twenty-four hours after transduction, the virus-containing medium was replaced with the second virus-containing supernatant. 48h after last transduction fibroblasts were harvested by trypsinization and replated in a 100mm dish seeded with 10^6 Nuff feeders. Next day medium was replaced with hES medium supplemented with 4ng/ml bFGF. The medium was changed every day for 30 days. After 10 days hES replaced by conditioned hES medium.

2.3.2 Using rAAV

rAAV vector particles were made by transient transfection of 293T using PEI, an adenoviral helper plasmid (HGT1), AAV pseudotyped vectors (2, 5 and 8) and transgene plasmids (expressing reprogramming factors and miRNA) as described in Davidoff *et al.*¹²⁴. Virus vectors were purified by the previously described ion exchange chromatography method¹²⁴.

Human fibroblasts were plated at $5-10 \times 10^4$ cells per well from 6 well plate and incubated overnight. For 4 days, cells were transduced sequentially with OCT4, SOX2, Nanog and Lin28 AAV at MOI of 2×10^6 . 48h after the last transduction, medium was changed to hES medium supplemented with 4ng/ml bFGF. The next day fibroblasts were harvested by trypsinization and replated in a 100mm dish on a Nuff feeder layer. The medium was changed every day for 30 days. After 10 days hES medium was replaced by conditioned hES medium. Similar protocol for AAV expressing miRNA 302

2.3.3 Using episomal plasmids

Proliferating human fibroblasts ready for transfection were harvested by trypsinization and 10^6 cells resuspended in 100 μ l nucleofector solution (Amaxa Nucleofector kit R - Lonza) that had been equilibrated to room temperature. Appropriate amounts of episomal plasmids added directly to resuspended cells: Yamanaka's plasmids (1 μ g of each of pCXLE hSK, pCXLE hUL and pCXLE shp53 hOCT4), Thomson's plasmids (7.3 μ g EBNA EN2L and 3.2 μ g EBNA ET2K) and miR302 plasmid (1 μ g). The cell suspension was gently mixed and transferred to a cuvette. Cell suspensions were transfected using program U-023 of Amaxa Nucleofector 2b machine. Immediately following transfection, cells were re-suspended in 10ml of pre warmed MEF medium and transferred to 0.1% gelatinised 100mm dish. Daily medium changes with MEF medium supplemented with 0.5mM sodium butyrate was started the next day. On day 7 post transfection, cells were harvested by trypsinization and 10^5 plated into 100mm dish with Nuff feeder cells in MEF medium. The next day, medium was replaced with hES medium with 10ng/ml of bFGF and 0.5mM sodium butyrate. By day 12 post

transfection, sodium butyrate treatment was stopped and conditioned hES medium supplemented with 10ng/ml of bFGF was used instead until day 30 post transfection.

2.3.4 Using episomal plasmids in feeder free conditions

10^6 fibroblasts were transfected with episomal plasmids as described above. However, at day 7 post transfection 2×10^5 viable cells were seeded into one well of a 6 well plate coated with matrigel/geltrex, recombinant vitronectin, laminin-521 or Synthemax-R in mTeSR1 or E8 medium supplemented with 0.5mM sodium butyrate. At day 12 post transfection, sodium butyrate treatment was stopped and cells continued to be cultured in mTeSR1 or E8 only until day 30 post transfection.

2.4 Pluripotency

2.4.1 Embryoid body formation

To demonstrate spontaneous *in vitro* differentiation, iPS cells were grown to confluency and harvested. Cells were re-suspended in hES medium without β -Fgf and transferred to non-tissue culture treated 6 well plates (Grenier Bio One). Medium was changed every 3-4 days. Day 12 embryoid bodies (EBs) were transferred to a 12 well plate coated with 0.1% gelatine and cultured for a further 12 days in fibroblast medium. Differentiated cells were harvested for RNA isolation.

2.4.2 Teratomas

For *in vivo* pluripotency, iPSC cells were grown to confluency and harvested. Cell pellets were re-suspended in 30% matrigel and 70% medium (mTESR1 or E8). Total of one confluent 6 well plate were used per animal. 6-8 week old NOD-SCID mice were anaesthetised by isoflurane inhalation and 50µl of the iPSC suspension was injected into each testis capsule. Analgesia (Carprofen, 5mg/kg) was administered intraperitoneally following surgery to minimise pain. After 10-12 weeks, animals were sacrificed by carbon dioxide asphyxiation and teratomas excised and fixed in HistoChoice® Tissue Fixative (SIAL). Histological processing and H&E staining were performed by either, The Research Department of Pathology, Faculty of Medical Sciences, University College London or The Research Department of Oncology, UCL Cancer Institute.

2.4.3 Scorecard

Pluripotency and trilineage differentiation potential was also assessed using the TaqMan® hPSC Scorecard™ kit 384w following manufacturer's instructions and run on a ViiA 7 system. Data analysis was performed using the cloud based TaqMan® hPSC Scorecard™ analysis software.

2.5 RNA isolation, RT-PCR and qPCR

MicroRNAs were isolated using mirPremier microRNA isolation kit from Sigma. Reverse Transcriptase PCR of microRNA was performed using MultiScrib Reverse Transcriptase from Invitrogen. Finally expression was analysed using miR302 TaqMan MicroRNA Assay. All procedures were done according to manufacturers' instructions.

RNA extraction was performed using RNeasy kit (Qiagen) or using Trizol method (Invitrogen). 1-3µg of RNA was used for subsequent reverse transcriptase reactions with

Superscript III first strand synthesis system (Life Technologies) as per manufacturer's instructions.

1 µl of cDNA was used per PCR reaction with GoTaq green mastermix (Promega). All procedures were done according to manufacturers' instructions.

cDNA samples diluted in appropriate volume in H₂O and 5 µl of cDNA was used per qPCR reaction with SYBR Green (Qiagen). All procedures were done according to manufacturers' instructions. QPCR data analyzed using the standard curve method for gene expression. GAPDH used as endogenous control.

List of primers used in Table 2 and Table 3

2.6 Phenotype analysis by flow cytometry

Cells were prepared as a single cell suspension in PBS/0,1% FBS and were labelled with the following fluorochrome-conjugated monoclonal antibodies (mABs): APC mouse anti-human CD184 (BD pharmingen) and ROR1 (Biolegend). Tra-1-60 (Santa Cruz), Tra-1-81 (Santa Cruz), SSEA5 (eBioscience) are non-conjugated mouse monoclonal IgM antibodies used in combination with an AlexaFluor®488 goat anti-mouse IgG, IgM (Life Technologies). Control staining with appropriate isotype controls was performed using APC mouse IgG1 (BD Bioscience and Biolegend) and AlexaFluor®488 goat anti-mouse secondary alone. For viability cells were stained with propidium iodide (PI, BD Bioscience) and analysed using FACS.

Samples were analysed by using either BD FACSVerse™ flow cytometer with BD FACSuite™ software or BD Accuri flow cytometer with BD Accuri C6 software.

2.7 Immunostaining

Cells were washed once with PBS and then fixed with 4% paraformaldehyde (PFA, Santa Cruz) for 20 minutes. Cells were washed 3 times with PBS and then stored at 4°C in PBS. Prior to immunostaining, cells were blocked for 30 minutes with 5% goat serum (Life technologies), 0.1% bovine serum albumin (Sigma) in PBS. If performing nuclear staining, cells were permeabilised with ice-cold 100% methanol for 5 minutes prior to blocking. Primary antibodies SSEA5, SOX17 (Sigma), HNF4a (Santa Cruz) and AFP (Santa Cruz) were used at a 1:100 dilution and used to stain cells at room temperature for one hour. Primary antibodies were washed off 3 times with PBS and then goat anti-mouse IgG/IgM Alexa 488 diluted 1:400 added. Secondary antibody was incubated at room temperature for 30 mins. Cells were washed 3 times with PBS and then DAPI (Cell Signalling Technology) added for 5-10 mins. Cells were washed 3 times with PBS and immediately visualised. For alkaline phosphatase staining, cells were stained with the alkaline phosphatase staining kit as per manufacturer's instructions (Stemgent or Millipore).

2.8 Hepatic differentiation

The *in vitro* differentiation protocol was similar to Chen *et al.* In brief, human iPS cells were used either as 70% confluent colonies or as single cells in mTeSR1. For single cell differentiation, cells were collected using accutase (Sigma) and plated at 2.6×10^5 cells/cm² in mTeSR1 medium supplemented with Y27632 ROCK inhibitor (CALBIOCHEM). For endodermal differentiation mTeSR1 medium was replaced with Roswell Park Memorial Institute (RPMI-1640, Sigma-Aldrich) with B27 (Invitrogen), 100 ng/mL activin A (Preprotech), 50 ng/mL Wnt3a (R&D Systems), and 10 ng/mL HGF (Life technologies) for 3 days, medium changed daily. During the next step, the culture medium was replaced with hepatic commitment medium: knockout/DMEM containing 20% knockout serum replacement, 1 mM L-glutamine, 1% NEA, 0.1 mM 2-mercaptoethanol, and 1% dimethyl sulfoxide. By the end of the hepatic commitment stage, cells were collected using accutase and plated at 2.1×10^5 cells/cm², cultured in Iscove's modified Dulbecco's medium (IMDM, Life Technologies) supplemented with 20 ng/mL oncostatin M (R&D systems), 0.5 µM

dexamethasone (Sigma), and 50 mg/mL Insulin-Transferrin-Selenium premix (ITS, Life Technologies).

All medium prepared under sterile conditions and filtered through a 0.22 mm pore-size filter (Merck Millipore).

2.9 Periodic Acid-Schiff (PAS) Stain for Glycogen

Cells were fixed in 4% PFA, and then permeabilized with ice cold methanol for 10 min. Samples were then oxidized in periodic acid solution (Merck Millipore) for 5 min, rinsed for 3 minutes in deionized water (dH₂O), treated with Schiff's reagent (Merck Millipore) for 15 min, and rinsed in dH₂O for 3 minutes. Samples were counterstained with hematoxylin solution (Merck Millipore) for 2 minutes and then rinsed in dH₂O for 3 minutes. Cells were incubated sequentially in ethanol 70%, 96% and 100% for 2min each. Lastly cells were incubated with Xylene (Merck Millipore) for 5 minutes and finally samples were observed under the light microscope.

2.10 ELISA Analysis

For the determination of Albumin production by Hepatocyte-like cells, cells and blood samples were collected and used in human serum albumin kit (Aviscera Bioscience).

2.11 Liver failure model

Male BALB/c nude mice, 8 weeks old and weighing 15–20 g, were used for our experiments. Fulminant hepatic failure was induced by Intra Peritoneal (IP) injection of CCl₄ (Sigma) 2.5 mL/kg body weight; 1:10 v/v in mineral oil. Solution kept on ice until IP injection in nude

mice using insulin syringes (TERUMO). Hepatic damage assessed by measurement of survival and histological analysis of organs.

For liver rescue, 24 hours after CCl₄ administration, mice were injected with Hepatocyte-like cells day 8 or Hepatocyte-like cells day 13. 1×10^6 cells/mouse diluted in PBS were administered via tail vein injection in a total volume of 100 μ l. In order to observe rescue of hepatic damage, measurement of survival and collection of organs for histological analysis was performed.

2.12 Isolation of mouse embryos

CD1 mice were mated, and conception was assessed by the presence of a coital plug with morning of the plug being scored as 0.5 days post coitum (E0.5). Pregnant females were sacrificed at various developmental time points and embryos were carefully removed by dissection.

Dissection of embryos was done under microscope with separation of the liver bud from the rest of the embryo when possible (termed “liver” and “embryos” respectively). Samples of all embryos from same plug were pulled together (except for E11.5 where individual embryos were enough for RNA extraction). Samples were placed in liquid nitrogen and stored at -80C until RNA isolation.

2.13 GEO datasets

Query GEO to obtain list of datasets fitting our settings (foetal mouse liver samples and human iPSC hepatic differentiation). Gene expression data was obtained and analysed using Affymetrix consoles: Expression Console (EC) and Transcriptome Analysis Console (TAC). Figures obtained from TAC both on hierarchical clustering and gene expression data. Affymetrix programmes downloaded from Affymetrix site.

2.14 Bioinformatics

Bioinformatics analysis of our RNA-seq data starts with aligning the reads to the reference sequence, the human genome (Ensembl build 70, accessed October 2014) using STAR, a splice-aware aligner. The advantage of aligning to the reference genome is the identification of novel features, if any. We have not found any such feature, which in respect to what we are looking for, is good since our genes of interest are already very well annotated in the human genome. Post-alignment analysis includes the quantification of gene or transcript isoform expression. The RNA-seq data are “analogous” and the quantification of expression is based on the one-to-one relationship between RNA fragments and reads (Figure 7). The quantification process requires a normalisation step that takes into consideration the length of the gene or transcript and the total number of reads mapped. Expression is reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Downstream analysis includes guided assembly of the transcriptome, where the assembly uses any existing transcript annotation as a guide¹²⁵ and differential expression analysis for the identification of statistically significant differences in expression levels between different conditions¹²⁶⁻¹²⁸.

2.14.1 RNA-sequencing Analysis

RNA samples were obtained as previously described (section 2.5). RNA sample preparation kit and paired-end sequenced to a length of 125 bp from 3 biological replicates for each of the four time points of day 0 (d0), day 4 (d4), day 8 (d8) and day 13 (d13) during hepatocyte differentiation of iPS cells (Figure 8). The RNA-seq reads were mapped to the transcript and cDNA sets of the Ensembl human genome (version 70) using STAR and Bowtie 1.0.1 allowing for multiple mapping between reads and transcripts. The resulting BAM files were merged using SAMTOOLS and made ready for further quantification analysis.

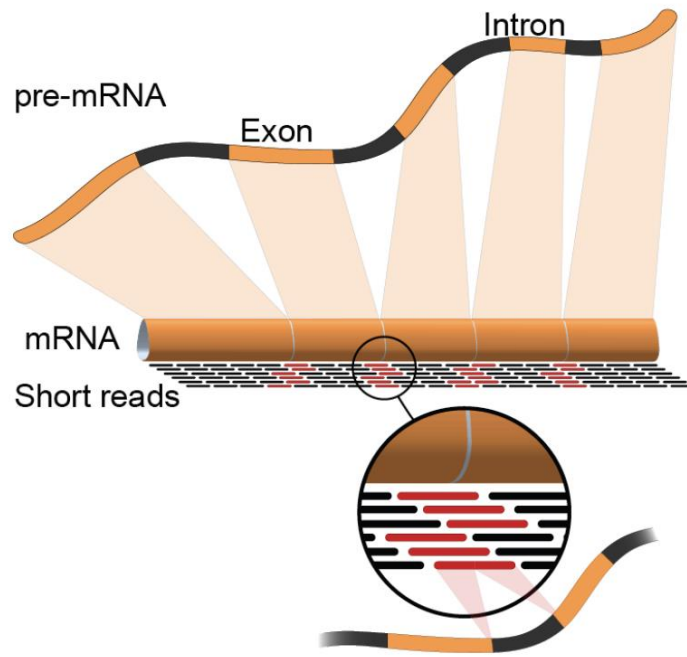


Figure 7 – One-to-one relationship between RNA fragments and reads. Adapted and edited from Garber *et al.* ¹²⁹



Figure 8 - Time-series RNA collection for analysis

2.14.2 Quantifying RNA-seq data

The resulting transcriptome alignments were used in MMSEQ gene expression algorithm to further quantitatively estimate transcript and gene expression levels. The MMSEQ gene expression analysis tool was used to estimate transcript expression levels (read counts). To perform a sample by sample comparison and estimate the expression from the RNA concentration in each microarray sample, the MMSEQ marginal posteriors for the set of Ensembl 70 transcripts mapped to by each Illumina BeadChip probe were collapsed.

Similarly, MMSEQ collapses marginal posteriors for the set of transcripts belonging each gene, thus providing gene-level expression estimates. Note that the expression estimates are roughly proportional to the RNA concentrations in each sample. The marginal posterior mean (\log_{μ}) and the posterior standard deviation (ds) of the mean corresponding to each transcript or set of transcripts (i.e., gene or probe) from the MMSEQ analysis was then used as the outcome in a Bayesian model selection algorithm implemented in the MMDIFF method and count data in the DNASEq2 software.

2.14.3 Rna-seq Analysis

DNASEq2 performs differential gene expression analysis based on the counts data from the MMSEQ results. The count data was normalized and quantified for the statistical inference of systematic changes between time point conditions (d0, d4, d8, d13). DNASEq2 was used to check if the results from MMDIFF¹³⁰ could be reproduced. Similarly, the \log_{μ} and sd from MMSEQ were quantified for the statistical inference of differential expression and fold change between two conditions in the regression based MMDIFF model which allows for comparison of complex experimental designs (Figure 9).

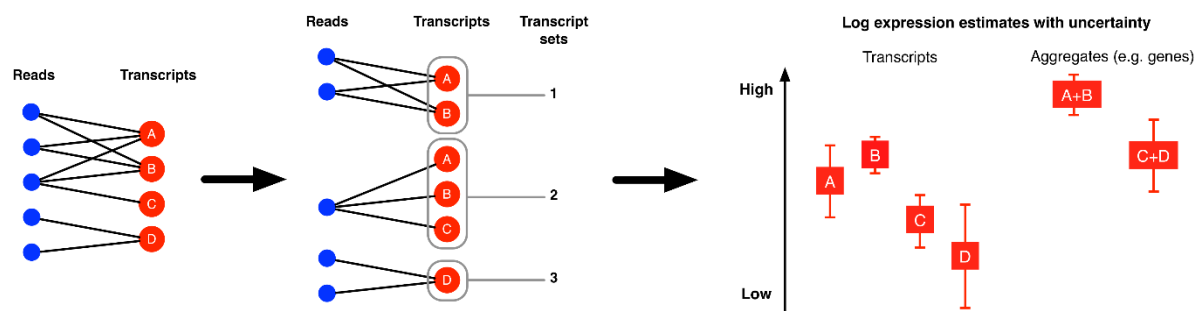


Figure 9 - Background of transcript quantification using MMSEQ and MMDIFF. Figure adapted from Turro *et al*¹³⁰.

The hypothesis of this study (based on what was observed in QPCR) was that few of the coagulation genes are expressed between d4 and d8 during hepatocyte differentiation and therefore the transcription factors required for the expression of these coagulation genes must be activated prior or along with these genes. Therefore, clustering and modeling the genes differentially expressed should result in identification of clusters within coagulation cascade based on their point of expression and expanding the same onto the entire data set would help identify the master regulators of the coagulation cascade. For differential analysis, the following design matrices were compared to quantify the differences in fold levels between d0 and d4, d4 and d8, d8 and d13 and d0d4 vs d8d13 (difference of difference of fold-change between two groups of two time points):

$$M^{(0)} = (1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1)^T$$

$$M^{(1)} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & -0.5 & -0.5 & -0.5 & -0.5 & -0.5 & -0.5 \end{pmatrix}^T$$

In another model, in order to quantitatively assess, if the log-fold change in expression between d0, d4, d8 and d13 differed with time, models were compared using the following matrix:

$$M^{(d0,d4,d8,d13)} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}^T$$

In both models, the rows correspond to d0, d4, d8 and d13 in consecutive set of three replicates and the prior distribution for the global intercept was fixed to $\alpha=0$. To identify differentially expressed genes between d0d4 and d8d13, the simple model assumed that the log-fold change between d0 and d4 was the same between d8 and d13, while the more complex model allowed the log-fold changes to differ. The prior probability of the more complex model being true was set to 10% ($p=0.1$). The posterior probability of the more complex model (Bayes Empirical factor) was used as the basis for preferring the more complex model (differential expression or difference of difference, respectively) to the simpler model in order to check if a set of genes were expressed between different time points. Only genes that were differentially expressed with a posterior probability < 0.5 were considered (note that posterior probability is different from p value).

2.14.4 Promoter Sequence Analysis for potential TFBS

The 5'UTR sequences of the 14 coagulation genes were obtained from Ensembl Genome Browser using in-house Perl scripts via Ensembl Perl API. These sequences containing transcription factor binding sites (TFBS) were aligned using pro-coffee which uses a substitution matrix estimated on TRANSFAC multiple sequence alignments of TFBS. The alignment was manually curated and studied for potential TFBS that are shared between the coagulation genes keeping in mind their evolution.

2.14.5 Statistics

All the statics were performed in R, unless otherwise stated. Where p-values have been compared between different groups, Q-values from False Discovery Rate analysis have been calculated.

2.14.6 DATA

In addition to the sequence and TF data mentioned above, we also built local databases of human genes with meta data that allows extracting, genes expressed in the liver. All the gene names and IDs (ENTREZ, ENSEMBL, HGNC) were obtained from Ensembl BioMart and stored in our local database for easy access.

2.14.7 Graphs

All the graphs have been generated in R and networks drawn using Cytoscape 3.0. Inkscape was used to edit and generate publication quality high-resolution graphics.

2.15 siRNA

HNF4a siRNA (HNF4a ON-TARGET plus SMART pool) and negative control siRNA (ON-TARGET plus non-targeting pool) were purchased from Dharmacon.

HUH7 and HepG2 cells were plated at suggested densities in D10 medium: DMEM supplemented with 10 % HF FBS and incubated overnight. On day 1, cells were transfected with 25nM of siRNA using Dharmafect reagent according to manufacturer's instructions

Cells were incubated for 36 hours and used for further experiments: viability (PI staining), gene expression and western blot.

2.16 Western Blotting

Whole cellular protein extracts were obtained by incubating cell pellets with RIPA buffer (Sigma) supplemented with protease inhibitor (Roche) for 5 minutes, spinning down at 8500g for 10 min and collecting supernatant. Use protein or store at -80C until use.

Protein extracts were run in a NuPAGE 10% Bis-Tris gel (novex) 125V, 35mA for 2h. Protein was transferred to a nitrocellulose membrane (Bio-Rad) for 1:30 hours at 25V, 125mA.

Membrane was blocked by incubation with 5% non-fat milk in TBST (10 mM Tris, pH 8.0, 150 mM NaCl, 0.5% Tween 20) for 60 min. After incubation the membrane was washed once with TBST and incubated with antibodies against HNF4a (Santa Cruz) at 1:1000 and GAPDH (Cell Signalling) at 1:2000 at 4 °C overnight. Membranes were washed three times for 10 min and incubated with a 1:3000 dilution of horseradish peroxidase-conjugated anti-mouse (Dako) or anti-rabbit (Dako) antibodies for 2 h. Blots were washed with TBST three times and developed with the ECL Western Blotting substrate (thermos scientific) according to the manufacturer's protocols

2.17 Primers

Table 2 – List of human Primers

Name	Sequence
GAPDH Fwd	GGCAACAATATCCACTTTACC
GAPDH Rv	GGAGTCAACGGATTTGGTCGT
OCT4 Fwd	GACAGGGGGAGGGGAGGAGC
OCT4 Rv	CTT CCC TCC AAC CAG TTG CCCAA AC
SOX2 Fwd	GGG AAA TGG GAG GGG TGC AAA AGA GG
SOX2 Rv	TTG CGT GAG TGT GGA TGG GAT TGG TG
Nanog Fwd	CAGCCCCGATTCTTTCCACCAG
Nanog Rv	CGGAAGATTCCCAGTCGGGTT
REX1 Fwd	CAGATCCTAAACAGCTCGCAG
REX1 Rv	GCGTACGCAAATTAAAGTCCA
SOX17 Fwd	CAG TGA CGA CCA GAG CCA GAC C
SOX17 Rv	CCA CGA CTT GCC CAG CAT CTT
CXCR4 Fwd	CCGCATCTGGAGAACCAGC
CXCR4 Rv	GGTGCAGCCTGTACTTGTCCG
Nestin Fwd	AGCGTTGGAACAGAGGTTGG
Nestin Rv	CCTCTGGGGTCCTAGGGAAT
Brachyury Fwd	ATGATGGAGGAACCCGGAGA
Brachyury Rv	ACTGCATCTTTCGGGACCTG
FoxA2 Fwd	CTACGCCAACATGAACTCCA
FoxA2 Rv	CGGTAGAAGGGGAAGAGGTC
HNF4a Fwd	CCA AGT ACA TCC CAG CTT TC
HNF4a Rv	TTG GCA TCT GGG TCA AAG
AFP Fwd	CCAAACAAAGGCAGCAACAG
AFP Rv	CAGACAATCCAGCACATCTCC
Albumin Fwd	AACCTCTTGTGGAAGAGCCT
Albumin Rv	GACCACGGATAGATAGTCTTCTG
TTR Fwd	CATGCAGAGGTGGTATTCACAG

TTR Rv	TTGGTTACATGAAATCCCATCCC
Fibrinogen A Fwd	AACCGTGATAATACCTACAACCGA
Fibrinogen A Rv	AATATCAATGTCCACCTCCAGTC
Fibrinogen B Fwd	AATATCCCAACTAACCTTCGTG
Fibrinogen B Rv	TTTCCTCACATTCTTTGCCA
Fibrinogen G Fwd	TGTGTTTCAGAAGAGACTTGATGG
Fibrinogen G Rv	ATATGTTAGGCGGTACTTGTCAG
Prothrombin Fwd	ACATAAGCCTGAAATCAACTCC
Prothrombin Rv	CGCTACAGTGACTTGATCCT
FV Fwd	CCATTTTCAGGACTTCTTGGG
FV Rv	TCCATTCATAGGTGTATTCTCGG
FVII Fwd	CACACCCACAGTTGAATATCC
FVII Rv	GCTCCATTACCAACAACAG
FVIII Fwd	CGCAAGATTTCTCCTAGAGTG
FVIII Rv	GACCTAGCAGACCCATCCAG
FIX Fwd	TCGGCCAAAGAGGTATAATTCAG
FIX Rv	GATCTCCATCAACATACTGCTTCC
FX Fwd	GGAAAGTCTGTTCATCCGCA
FX Rv	GCCTTCGAATCCTTCTAAACAG
FXI Fwd	TTACACAGATTCTCAACGACCC
FXI Rv	CCTACCAGATGCCAGACCTC
FXII Fwd	AGTCAACACTTTTCGATTCCACC
FXII Rv	GGTCTTTCACTTTCTTGGGCT
FXIIIa Fwd	CGAAACCCAGAAACAGACAC
FXIIIa Rv	ATGCCATCTTCAAACCTGACCA
FXIIIb Fwd	AACTCTATGCAGAAGAGAAACCC
FXIIIb Rv	GTTTGCTCTTCTTGTCTTCCA

Table 3 - List of mouse Primers

Name	Sequence
GAPDH Fwd	TGG AGA GCC CGC TCA GAC CC
GAPDH Rv	GGA TTG GGT GTC CCT GCG CC
Fibrinogen B Fwd	TCTGGGAAAGAGTGTGAGGA
Fibrinogen B Rv	GCCAAGCCAATATTCACCTG
Fibrinogen G Fwd	GTAGTTTCTGCCCAACCACC
Fibrinogen G Rv	TCCTGTAAATACCGAATGCTTGTC
Prothrombin Fwd	ACAAGCCTGAAATCAACTCC
Prothrombin Rv	TGTTGTACAGCCAAATTCC
FV Fwd	CAAACGCCATTTCTCATCATAGAC
FV Rv	CTCCCAATAAGTCAGATATTCCGA
FVII Fwd	ATGAGGACTACACGCTACAG
FVII Rv	GAAGTCATGTTACCCATCAC
FVIII Fwd	GGCACTCAGAAACAAACGAC
FVIII Rv	CATCCAATCAGACCTGGAAGAG
FIX Fwd	GCTCTCATCACCATCTTCCT
FIX Rv	ACACTGATCTCCATCAACATACTG
FX Fwd	CTGCTCAACCTCAACGAGAC
FX Rv	GTGTTCCGATCACCTACCCT
FXI Fwd	AGTGTCCCAGTATTCTGCCA
FXI Rv	GAGTATCCAGAGATGCCTCCC
FXII Fwd	TTTCCCTTTCAGTACCACCGT
FXII Rv	GGCTCAAAGCATTTCTCTTTCTG
FXIIIa Fwd	AGGAAGTACCCAGAGGCACA
FXIIIa Rv	CCCTCTGCGGACAATCAACT
FXIIIb Fwd	CCAAGGTGCTACAAGAAATGTC
FXIIIb Rv	GCATGTGTAAGCCACTATGTC

Chapter 3 Integration free reprogramming method

3.1 Introduction

Pluripotent stem cells have unlimited self-renewal and the capacity to differentiate into all cell types of the three germ layers: endoderm, mesoderm and ectoderm. Therefore, they have enormous potential for regenerative medicine as they can serve as an endless source of cells for the repair, replacement or regeneration of damaged or diseased tissues. Furthermore they can serve as an important tool for the study of disease mechanisms as well as drug testing. ESCs are such cells and great promise surrounded them after their discovery.

In 1998, James Thomson derived the first human ESCs from the ICM of the blastocyst¹³¹. Since then, many protocols have been developed showing that ESCs can be differentiated into cells of various lineages including neurons, cardiomyocytes, blood progenitors, hepatocytes and retinal precursors¹³². However, the use of human ESCs has been hindered by the ethical concerns over the destruction of embryos. Another drawback is the difficulty in generating patient- or disease-specific ESCs due to the logistical difficulties in obtaining embryos.

In 2006 a breakthrough was achieved by Kazutoshi Takahashi and Shinya Yamanaka when the first iPSCs were created from murine fibroblasts¹⁴. These iPSCs were obtained by the over-expression of four transcription factors: Oct4, Sox2, Klf4 and c-Myc by retrovirus-mediated transduction of fibroblasts¹⁴. These 4 factors were chosen from a list of 24 genes that had been shown to be crucial for maintenance of pluripotent stem cell identity and proved to be sufficient for iPSC formation¹⁴. A year later, the same group generated human iPSCs using the same 4 transcription factors¹³³. The mechanisms by which the reprogramming occurs are poorly understood but it was clear that the expression of the exogenous genes was only transiently required as upon reprogramming of pluripotent state, the endogenous genes Oct4, Sox2 and Nanog are reactivated, while the transgenes are

silenced. The generated iPSCs are similar to ESC in morphology, growth properties and the expression of phenotypic markers. Teratoma formation was used to assess the pluripotency of the iPSCs and confirmed the ability of these cells to differentiate into all three germ layers *in vivo*. Despite the similarity between ESCs and iPSCs, these are not identical with small differences in gene expression and DNA methylation patterns^{82, 85}. Interestingly, the same has been observed between different ESC lines¹³⁴ and some studies show that is difficult to distinguish between iPSCs and ESCs^{90, 91}, thus the implications of these minor differences remain unclear.

Soon after Yamanaka's discovery, Thomson revealed a different combination of reprogramming factors, Oct4, Sox2, Nanog and Lin-28 (OSLN) which were also capable of generating iPSCs¹³⁵. This combination lacked the known oncogene, c-Myc and was therefore deemed to be safer. Since then, many other combinations of reprogramming factors have been described. Some of these combinations include the use of chemicals and/or small molecules such as valproic acid and RepSox which can replace some of the transcription factors^{20 136}. It has also been extensively demonstrated that the somatic cell type can greatly influence the efficiency of iPSC derivation, for instance, CD34+ cord blood stem cells have been found to be one of most easily reprogrammed¹⁵. The multitude of factors and molecules found to influence in the reprogramming process shows the complexity and highly interconnected pathways necessary for the establishment of the pluripotent state (Figure 10).

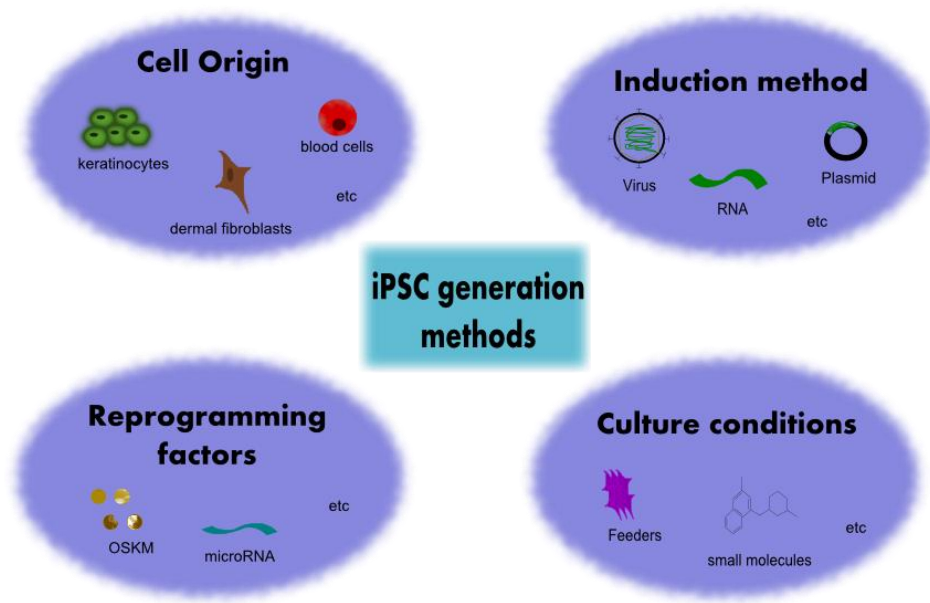


Figure 10 - iPSC generation methods. Factors affecting iPSC reprogramming can be categorized into four groups (1) the origin of the cells (2) the induction method (3) the combination of reprogramming factors and (4) the culture conditions

With the realization that pluripotency could be achieved using a wide range of methods, people started focusing their efforts on creating ‘safer’ iPSCs in anticipation of their use for cellular therapies. This was done by decreasing their tumorigenic potential, mainly working on two requirements: avoiding the use of genome-integrating methods and removal of oncogenes in the reprogramming factor cocktail. By using integrative methods we risk the activation of oncogenes and/or disruption of tumour suppressing genes which would thwart their use in a clinical setting. By using reprogramming factors that include oncogenes, such as c-Myc, there is always the risk of their re-activation.

Many methods for generating iPSCs without the use of integrating vectors or oncogenes have since been developed. However, some of these methods such as the Cre-deletable viral vectors and PB transposons are not strictly integration free as they can leave genomic scars behind¹³⁷. While other methods, such as protein delivery, avoid the introduction of exogenous genetic material, they are weighed down by low efficiency and slow kinetics of

reprogramming¹³⁸. The advantages and disadvantages of some of the methods are outlined in Table 4.

Despite the work done on developing the different methodologies, not much work has been invested on comparing them side by side – the only way to truly access which one is the best in terms of efficiency. With the wide range of methods available it is impossible to perform a systematic comparison between them all. With the goal of clinical translation, we decided to choose the methods that showed the greatest potential amongst the integration-free methods: viral delivery, episomal plasmid delivery and RNA delivery and compared this with the original retrovirus standard.

Table 4 – Summary of advantages and disadvantages of different reprogramming method

Reprogramming method	Advantages	disadvantages
Retrovirus	Efficient and stable	Genome integration; transgenes present in the genome which may not be properly silenced
Lentivirus	Efficient and stable	Genome integration; residual expression of the transgenes
Adenovirus	Rare genomic integration and vector lost by dilution	Slow and inefficient
Episomal non-replicative vectors	Rare genomic integration	Multiple transfections required; Slow and inefficient;
oriP/EBNA episomal vectors	Rare genomic integration	Slow and low efficiency
Mini circle	Rare genomic integration	Slow and low efficiency; labour-intensive
RNA	No genomic integration; higher efficiency than retrovirus; No need to screen numerous colonies	Multiple transfections required; Expensive; labour-intensive
Proteins	No genomic integration; No need to screen numerous colonies	Slow and inefficient; Expensive

3.2 Viral delivery systems for reprogramming

Two different viral approaches have been shown to produce integration free iPSC: adenovirus and Sendai virus^{45,47}. The efficiency of the adenovirus approach was found to be 3 orders of magnitude lower than the retroviral approach (0.0001%-0.0018%) and hepatocytes were the only cell type that could be successfully reprogrammed without the help of a complementary vector⁴⁵. On the other hand, F-deficient Sendai virus could reprogram both fibroblasts and T-cells to a pluripotent state while viral RNA was not detected in the resulting iPSCs. Despite the use of Sendai viral vectors with a single gene defect, safety and regulatory concerns remain since F-deficiency only decreases the likelihood of viral self-replication but does not eliminate it. A further hindrance for the use of sendai virus in a clinical setting, is their high cost relative to other methods, which could eventually limit the availability of this technology to patients⁴⁷.

3.2.1 AAV reprogramming

A recombinant adeno-associated virus (rAAV) vector mediated reprogramming of human somatic cells, could represent a new integration free viral method of reprogramming since AAV genomes normally don't integrate into the host's genome. rAAV offers a "hit and run" strategy for reprogramming somatic cells as they are kept episomally and are lost over time with cellular division. Furthermore, AAV vectors are attractive for generating clinical grade iPSCs as they have a great safety record in humans^{139 140}.

The efficiency of rAAV transduction to fibroblasts was first optimised to ensure that the reprogramming factors could be delivered into the cells. As certain AAV serotypes are known to have tropism for different cell types, we first sought to determine which serotype would be best for transducing human fibroblasts. We generated rAAV containing the green fluorescence protein (GFP) transgene with 3 different serotypes: AAV2, AAV5 and AAV8. AAV8 showed the lowest transduction efficiency (24%) as expected, due to its tropism for

liver cells while AAV2 and AAV5 showed similar GFP expression ($74\pm 8\%$ and $72.5\pm 4\%$ respectively) (see Figure 11). Due to the low yields of AAV2 during virus production, AAV5 was chosen for generating rAAV containing the reprogramming factors. As we aimed to use an integration-free reprogramming method and avoid the use of oncogenes, we decided to go with Thomson's set of transcription factors¹³⁵ (OSLN) as these are deemed to be safer than Yamanaka's original set (OSKM).

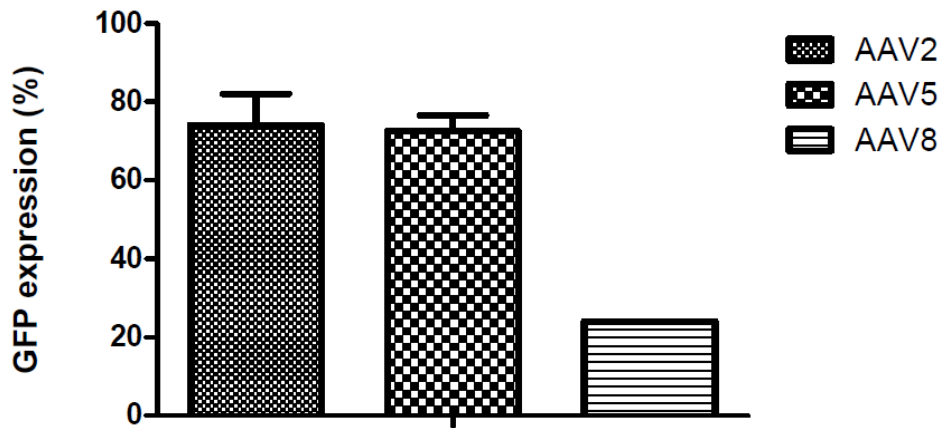


Figure 11 –Transduction efficiency of IMR90 fibroblasts using different AAV serotypes. Transduction of IMR90 fibroblasts with AAV serotypes 2, 5 and 8 expressing GFP under CMV promoter. AAV virus were added to IMR90 cells and incubated for 48h before accessing transduction efficiency. Transduction efficiency as percentage of GFP positive cells as detected by flow cytometry..

A limitation of rAAV is the packaging capacity (approximately 4.7kb). Due to the size of the transcription factors (approximately 1000 base pairs or more), it was necessary to generate 4 different rAAV particles for each of the 4 transcription factors (OSLN). rAAV expressing GFP was also generated to allow the transduction efficiency to be measured. As a control, we first demonstrated that both the BJ cell line (derived from foreskin of newborn human) and a primary patient cell line (RPD1) could be readily transduced with GFP retrovirus (BJ: $97.8\pm 0.318\%$; RPD1: $98.8\pm 0.084\%$). The same cell lines could also be transduced with rAAV GFP (BJ: $85.5\pm 0.707\%$; RPD1: $79.5\pm 0.707\%$) at similar efficiencies, as shown in Figure 12.

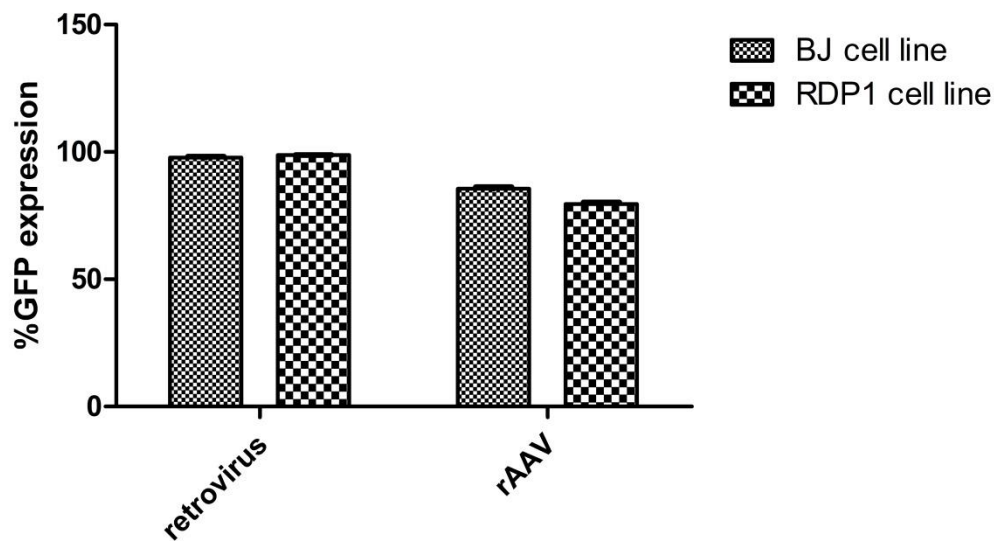


Figure 12 - Comparison of GFP transduction efficiency of fibroblasts using retrovirus and rAAV. Transduction of BJ and RDP1 cell lines using retrovirus and rAAV expressing GFP under CMV. AAV virus were added to IMR90 cells and incubated for 48h before accessing transduction efficiency. Transduction efficiency as percentage of GFP positive cells as detected by flow cytometry.

The lack of knowledge regarding the process of reprogramming with AAV led us to decide to add the virus to the cells under two different conditions: all factors together and sequential addition of each factor. Addition of all the factors mirrors the original retroviral reprogramming protocol while the sequential introduction of factors would work as an attempt to increase the levels of expression of the individual factors, as later corroborated by Liu *et al.*¹⁴¹.

While 4 factor transduction of BJ and RDP1 cells with retroviruses resulted in 112 and 49 TRA-1-81 positive iPS cell colonies, respectively, no successful reprogramming was achieved using rAAV in both conditions (n=2). A possible reason for this may be the competitive nature of rAAV particles. This competitive nature results from the fact that single-stranded rAAV with sense (plus) and anti-sense (minus) orientation are packaged equally well and upon transduction these form transcriptionally active dimers by the

annealing of single-stranded (ss) molecules with opposing polarities. This is an inefficient process under normal conditions and the presence of four different ssAAV cassettes in the same cell raises the possibility of reducing the efficiency by which double stranded AAV expression cassettes are formed. Serial transduction of each factor potentially offers a level of protection from this mechanism but it may not be enough. Alternatively, it is possible that the 4 factors were not expressed in the target cells at the right time and levels required to promote reprogramming. Therefore, it may be advantageous to reprogram using only a single rAAV cassette that dispenses the need for 4 different rAAV vectors. Due to the limiting packaging capacity of rAAV, it isn't possible to construct a polycistronic vector containing all 4 factors. Under this premise we thought of using a single cassette expressing microRNA (miR) 302 cluster for AAV reprogramming.

In 2011, microRNAs were also shown to be capable of inducing pluripotency²³. MicroRNAs are small single stranded RNA molecules of about 22 base pairs that bind to complementary sequences in mRNA and regulate expression by gene silencing. These regulators are present in the human genome and are thought to target around 60% of the genome¹⁴². Due to the redundancy of each molecule, a single miRNA can have up to hundreds of targets. MicroRNAs are known to regulate development and differentiation throughout different tissues^{143 144}. In ESCs, the expression of specific miRNAs is critical in the control of pluripotency-related genes^{145 146}. Of these pluripotency associated miRNAs, the miRNA 302 cluster (miR302) showed concomitant expression with OCT4 through development and in the same tissues, as well as being a direct target of both OCT4 and SOX2 – two of the critical factors required for iPSC reprogramming¹⁴⁷. This cluster consists of 5 different miRNAs which are transcribed together, four of which belong to the miR302 family and are highly homologous: mir-302b, mir302c, mir302a, mir302d and mir367¹⁴⁸. Reprogramming into an iPSC state was achieved directly using this miR302 cluster alone at a higher efficiency than with the standard OSKM retroviral method²³. The use of molecules which lacks any further processing (protein translation) might be one of the underlying reasons of the higher efficiency observed. The multitude of miRNA targets may also allow for faster and more coordinated reprogramming. However, in this report, the miR302 cluster was expressed in lentiviral vectors which are not the ideal methodology as these vectors integrate within the host's genome²³.

We exploited this finding by attempting to reprogram our cells using a single rAAV particle expressing the miR302 cluster alone. However, no iPSC colonies could be generated using this method. Expression of the mir302 cluster could be detected in transient transfections of 293T cells with the miR302 vector plasmid but when the same 293T cells were transduced with the rAAV-miR302, the micro RNA could not be detected. The reasons for this are not clear but it is possible that interaction of the miR302 with the secondary structures of the inverted terminal repeats (ITR) within the genome of the single stranded proviral DNA could have interfered with its expression.

During this period, Weltner and colleagues reported that rAAV of the 4 factors could be used to induce pluripotency in MEFs but not in human fibroblasts¹⁴⁹. In this study, high viral titers were used to ensure adequate transduction which may be the reason behind the increased incidence of rAAV vector integration into the host genome. Furthermore, the authors found that the integrated rAAV transgenes were not silenced over time. Regardless of the underlying reason, this finding excludes rAAV as an integration-free approach. The low reprogramming efficiencies of MEFs (0.001-0.09%) coupled with the high rates of rAAV vector genome integration made the rAAV method unsuitable for generating clinical grade iPS cells and so the rAAV approach was abandoned.

3.3 Episomal reprogramming

Episomal plasmid vectors carrying the OSKM factors were first used to successfully induce pluripotency in MEFs but were unable to produce any iPSCs from human cells^{150 49 50}. Possible reasons behind this failure are the low transfection efficiency of large plasmids, dilution of vectors in actively proliferating cells, or even the silencing of prokaryotic sequences contained in these vectors, which leads to a down regulation of the transcription factors. The timing of this down regulation is critical as human cells require a longer expression of OSKM than MEFs to reach pluripotency.

Thomson was able to circumvent the problem of episomal dilution by using the oriP/EBNA1 episomal vector⁵¹. These plasmids can be transfected without the need for viral packaging, are maintained episomally throughout cell division and without selection are lost at ~5% per cell generation, thereby allowing removal of the vectors but still giving it enough time for the reprogramming process. In this way, Thomson was able to generate iPSC from human fibroblasts by introducing a combination of 7 transcription factors: OCT4, SOX2, NANOG, LIN28, c-MYC, SV40 T large antigen and KLF4⁵¹. Although representing the first successful case of episomal derived human iPSC, the efficiency of the process was extremely low (3 to 6×10^{-6} %).

To try and improve the efficiency of episomal reprogramming, Yamanaka combined Thomson's findings with two of his own: p53 suppression greatly enhances iPSC formation and L-Myc is more potent and specific than c-Myc during human iPSC generation⁵². In this way Yamanaka created his own set of oriP/EBNA1 vectors with 6 transcription factors (OCT4, SOX2, NANOG, LIN28, L-MYC and KLF4) and a small hairpin (sh) for p53. By introducing these 2 changes Yamanaka reported a higher efficiency than Thomson, producing integration free iPSCs at an efficiency suitable for stem cell therapy⁵². Even though Yamanaka showed a higher reprogramming efficiency, the differences between labs (from the technician, media and methods utilised) can affect the efficacy of the process, and so a side by side comparison in the same lab will be able to eliminate these from the equation. For

this reason we decided to first compare the Thomson episomal plasmids (encoding Oct4, Sox2, Klf4, c-Myc, Nanog, Lin28 and SV40 large T antigen) with the Yamanaka episomal plasmids (encoding Oct4, Sox2, Lin28, L-Myc and p53 shRNA) in order to confirm what was shown in the literature.

The method for generating integration free iPSCs using episomal plasmids is simple, involving only a single transfection and thus has great potential for clinical translation. However, the protocol requires further optimization as the reported reprogramming efficiencies are relatively low compared with the retrovirus standard^{51, 52}. For episomal reprogramming, the same stock and passage of the BJ fibroblast cell line used in the previous rAAV experiments were used here. The transfection efficiency of BJ fibroblasts using plasmids as measured by the pCXLE GFP construct was approximately $36\pm 8.1\%$ (mean \pm SEM, n=4) at 24 hours after electroporation (see Figure 13).

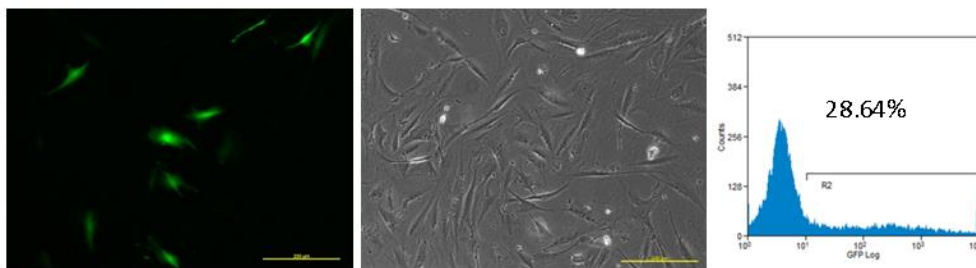


Figure 13 - Transfection efficiency of human fibroblasts. GFP plasmid transfection of BJ fibroblasts analysed by immunofluorescence and FACs. Episomal plasmid (1 μ) added directly to cells resuspended in 100 μ l nucleofector solution (Amaca Nuclofactor Kit R). Transfection efficiency as percentage of GFP positive cells as detected by flow cytometry. Scale bar is 200 μ m.

Morphology changes in fibroblasts were observed in induced cells with the Thomson plasmids as early as 7 days post transfection (see Figure 15A), while for the Yamanaka plasmids, changes were first observed at days 10-12. However, by days 25-30 more iPS-like colonies emerged from Yamanaka plasmids, with better morphological features than with the Thomson plasmids (Figure 15B and Figure 15C). *Bona fide* iPSC colonies were identified by TRA-1-81 live staining and 43 ± 29 colonies were observed for Yamanaka plasmids compared

with only 10.5 ± 2.5 positive colonies for Thomson plasmids ($n=2$). After taking into consideration the transfection efficiency, the computed reprogramming efficiency of the Yamanaka plasmid combination was found to be over 10 fold greater ($0.012 \pm 0.008\%$ Yamanaka vs $0.003 \pm 0.001\%$ Thomson) at generating iPS cells and was therefore used in all subsequent experiments. Reprogramming using patient fibroblast lines showed similar efficiencies to the BJ line (Supplementary figure 1).

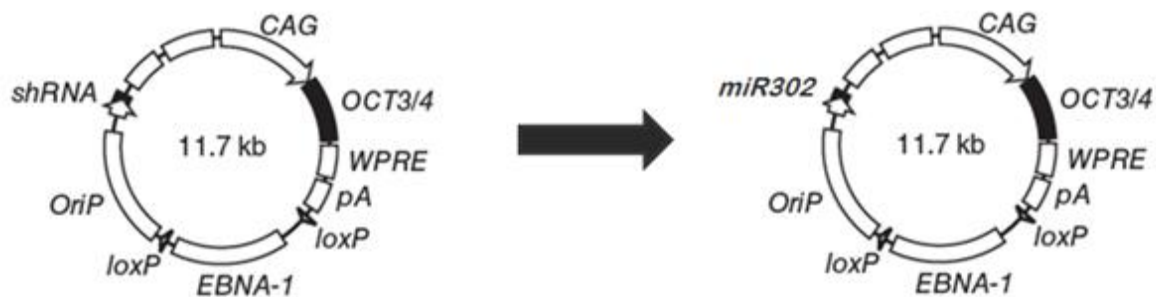


Figure 14 - Schematic of the cloning of the miR302 cluster into the pCXLE hOct4-shp53 plasmid

As with the rAAV strategy we decided to implement the miR302 cluster in the episomal plasmid approach to see if we could improve the efficiency observed with the Yamanaka plasmids. Therefore, the miR302 cluster was cloned into one of the episomal plasmids, pCXLE-hOct4-shp53, specifically the miR302 sequence replaced the shp53 sequences in this plasmid (see Figure 14) due to the similar pathways between miRNAs and sh RNAs.

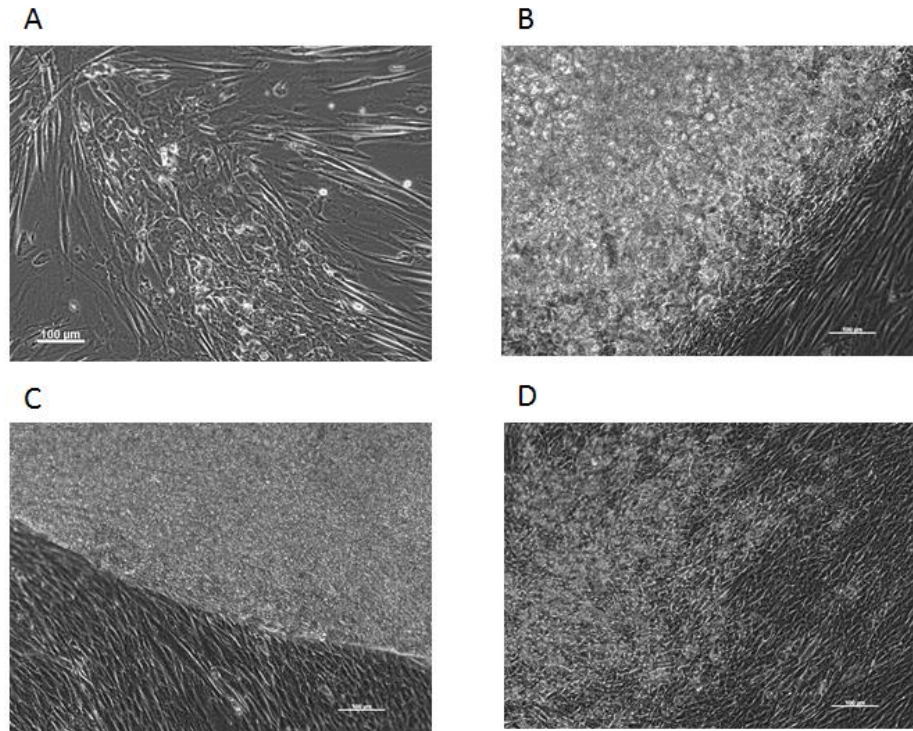


Figure 15 - Reprogramming with Yamanaka's, Thomson's and miR302 vectors. Phase contrast pictures of BJ fibroblasts reprogrammed using episomal plasmids A. BJ fibroblasts transfected with Thomson's vectors at day 7. B. BJ fibroblasts transfected with Thomson's vectors at day 30. C. BJ fibroblasts transfected with Yamanaka's vectors at day 30. D. BJ fibroblasts transfected with miR302 vector at day 30.

Transfections of BJ fibroblasts were performed with Yamanaka's episomal plasmids and the new miR302 plasmids to allow a side-by-side comparison. Some conformational changes could be observed in fibroblasts transfected with miR302 plasmid although no fully reprogrammed cells were present (see Figure 15D). To determine the reason behind this observation, BJ fibroblasts were transfected with the new pCXLE-hOct4-miR302 plasmid and its miR-302 expression level was compared with the H1 hESC line and un-transfected BJ fibroblasts. miR302 expression could be detected in transfected cells but levels were 3 orders lower than in hESCs (see Figure 16). The low expression levels reported for miR302 are the most likely reason for the absence of colonies as it has been shown that for efficient reprogramming of hair follicles, expression of miR302 needs to be at least 1.3 fold higher than in hESCs¹⁵¹. Even though miR302 can indeed achieve reprogramming, its use is dependent on the expression levels achieved. The delivery system can also impact on the

efficiency, as observed when the use of synthetic mature miRNAs¹⁵² for reprogramming did not match the levels observed with lentivirus.

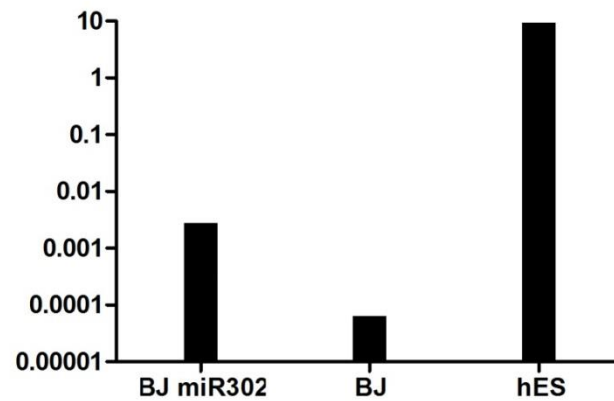


Figure 16 - Comparison of miR302 expression levels. miR302 expression levels in H1 ES cell line, BJ fibroblasts and BJ fibroblast transfected with pCXLE hOCT4 miR302 detected by qPCR. Expression levels normalized to RNU48 endogenous control.

The use of episomal plasmids to drive expression of the miR302 cluster is probably insufficient to reach the levels required to have a positive effect on the reprogramming process. While shp53 cannot by itself reprogram somatic cells back to a pluripotent state, its presence in the episomal plasmids improves the reprogramming efficiency more effectively than when the miR302 cluster was added. Based on this, the original Yamanaka episomal plasmids were adopted for all future experiments.

3.4 RNA reprogramming

The direct delivery of synthetic mRNAs into the cells completely avoids the use of DNA or viral vectors and is therefore ideal for clinical applications. This method requires serial daily transfections due to the short half-life of the mRNA molecules but was able to achieve the fastest and most efficient reprogramming reported so far⁵⁶.

For Warren *et al.* to be able to reprogram somatic cells using mRNA without causing cellular toxicity, they had to engineer the mRNA molecules in order for these to avoid the host cell's innate immune defense mechanisms, defenses that are in place for protection against RNA viruses⁵⁶. This was done by generating synthetic mRNA molecules with a phosphatase cap and modified ribonucleotides. These modifications coupled with supplementation of the interferon inhibitor, B18R in the medium attenuated the innate antiviral response and allowed high cell viability. Despite the high cost in production of the synthetic mRNAs and the technically demanding work required, the use of mature mRNA is highly appealing as it allows the step of mRNA transcription to be skipped, making this the possible reason behind the faster reprogramming time observed.

As observed by others, when reprogramming the BJ cell line using the mRNA methodology we observed the highest efficiency of reprogramming. The high transfection efficiency, approaching 100% (Figure 17B), greatly impacts the efficiency as the internalization of the mRNA cocktail in all the cells enhances the chances that all reprogramming factors will be present inside the cells. Morphological changes could be seen in the culture as soon as 4 days into the reprogramming, and colonies began to emerge after the 12 days of transfection, significantly earlier than observed by other methods (Figure 17A).

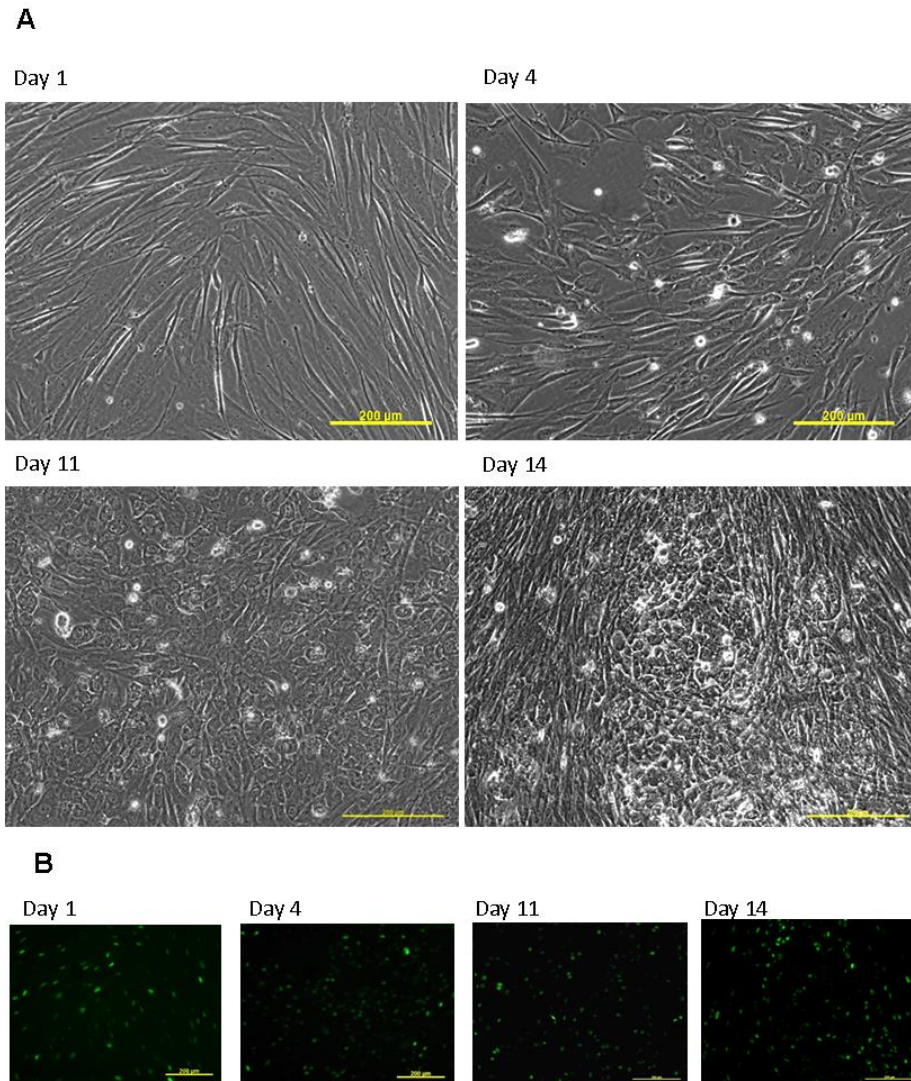


Figure 17 – mRNA reprogramming of BJ fibroblast line A. Phase contrast pictures of BJ fibroblasts transfected with reprogramming mRNAs at days 1, 4, 11 and 14 B. Immunofluorescence pictures of BJ fibroblasts transfected with GFP mRNAs at days 1, 4, 11 and 14.

From analysis of the culture at day 18, TRA-1-81 staining revealed a reprogramming efficiency of 1.89% - the highest observed between all the different methods used, including the original retroviral approach. Despite the promising results obtained from BJ cell line, when using primary patient fibroblast lines, no successful reprogramming was observed (Figure 18). This observation is not unique to our case as others have reported problems when using mRNA to reprogram patient fibroblast samples¹⁵³. Furthermore, no successful

reprogramming has been reported when using blood cells. The BJ line, derived from newborn foreskin, might be more susceptible to reprogramming, as it has been shown that the age of the patient can influence the efficiency of reprogramming¹⁵⁴.

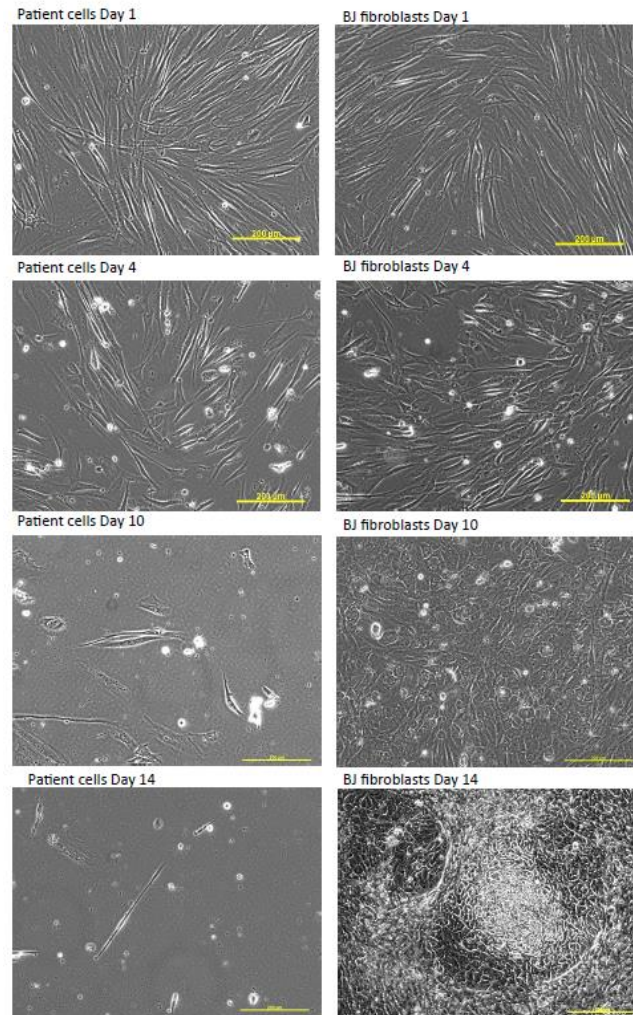


Figure 18 - Comparison of mRNA reprogramming of patient Vs BJ fibroblasts line. Inefficacy of mRNA to reprogram patient lines. Phase contrast pictures of mRNA reprogramming of patient line and BJ line at days 1, 5, 10 and 14.

As our choice of reprogramming methodology is linked to a practical application, the ineffectiveness of mRNAs with patient lines renders this technology obsolete. If this technology is to be used in regenerative medicine, we need a method capable of reprogramming cells from patients. Consequently we chose the episomal method of

reprogramming as, from all the integration-free methods, it was the only that was able to meet the required criteria: integration-free, efficient and able to effectively reprogram patient lines.

3.5 Characterization of the different lines

Several methods are available for reprogramming cells into a pluripotent state, and we conducted a side by side comparison of some that could be considered safe as they carry no or little risk of integration into the genome of the host cell. In order to identify the emergence of iPSC colonies among the fibroblast population, morphological changes are used as a crude evaluation of efficiency, while TRA-1-81 staining allowed for a more precise characterization. Based on this, several colonies were picked and expanded for further characterization.

Nevertheless, the premises on which the lines were established do not guarantee a true pluripotent state. During the reprogramming process cells undergo a series of changes, as a result of the introduction of exogenous factors. The process is still not completely understood and for that reason variability exists among different clones. Multiple lines can be established from one round of reprogramming but not all lines will present equal genetic, epigenetic or phenotypic characteristics. It is important to pick more than one clone and evaluate them to ensure their pluripotency, avoiding any potential issues with data misinterpretation at later stages.

The usual tests performed to characterize iPSC line include morphological analysis, *in vitro* differentiation by embryoid body (EB) formation, teratoma formation by injection of iPSC into immunodeficient animals, karyotypic analysis, expression of pluripotency markers such as OCT4, Sox2, Nanog, SSEA4, Tra-1-60, Tra-1-81 and integration status of reprogramming factors.

In frame with what is routinely done, we performed a series of tests for quality controls of the different lines established using the different reprogramming methods. Comparing iPSC established using different reprogramming methods can also allow to see if there is any underlying abnormality that is method specific.

All iPSC lines derived from BJ fibroblasts using retrovirus, episomal and mRNA showed the typical hES-like morphology with large nucleus to cytoplasm ratio, highly visible nucleoli and growth as compact colonies (Figure 19A)¹⁵⁵. Looking at pluripotent markers, both at the gene and protein level, we could observe the presence at similar levels to the control line H1 (hESC line) (Figure 19B and C). We also showed that the parental BJ fibroblasts were negative for pluripotent markers such as Nanog, Rex1 and Sox2 (Figure 19C).

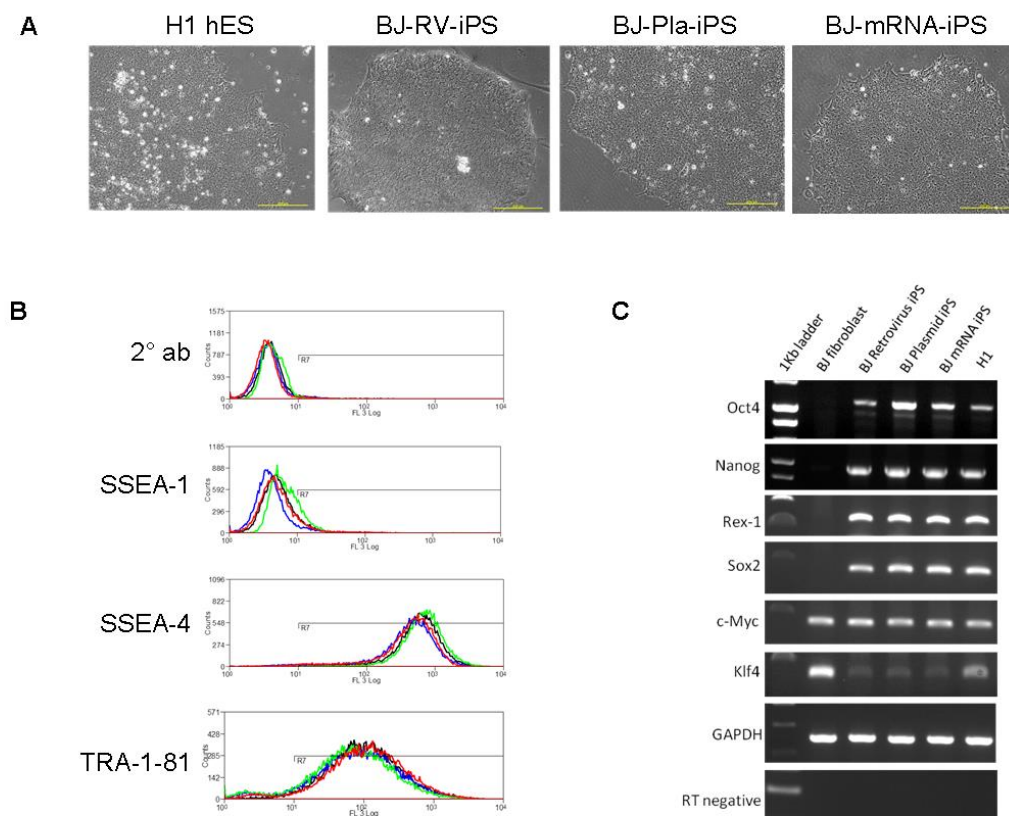


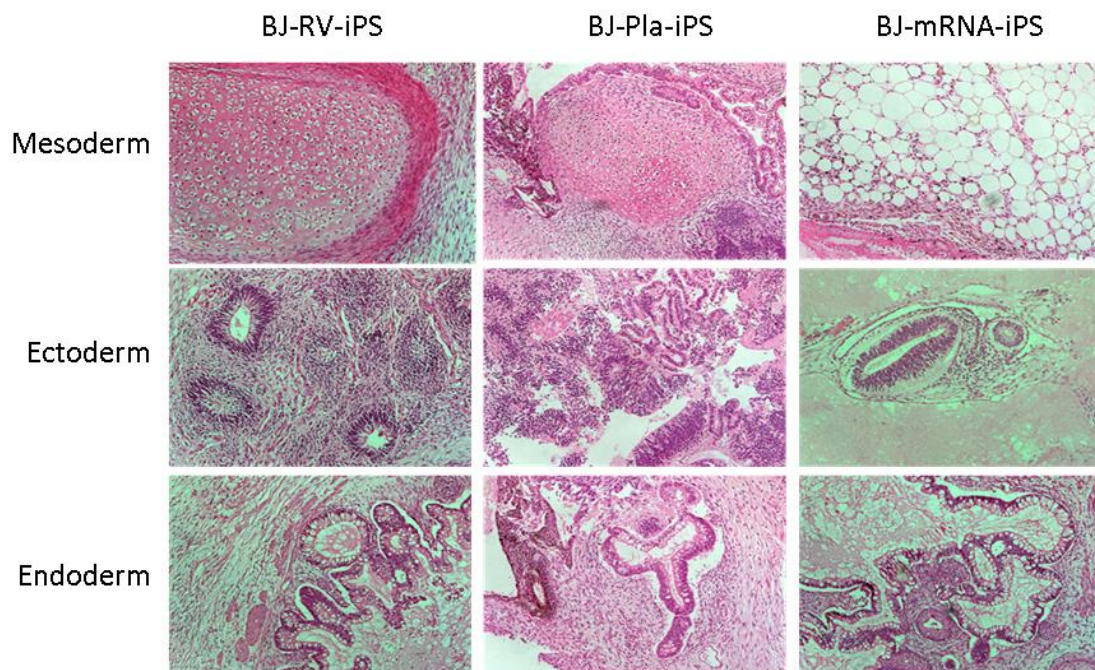
Figure 19 – Characterization of established iPSC lines derived from BJ fibroblasts using retrovirus, plasmids and RNA. A. Representative phase images of established iPCS and H1 cell lines. B. Flow cytometry analysis with pluripotency markers SSEA-4 and TRA-1-81. SSEA-1 is a negative marker of human pluripotent stem cells. Green line denotes H1, red line denotes BJ-RV-iPS, blue line denotes BJ-Pla-iPS and black line denotes BJ-mRNA-iPS. C. RT-PCR analysis for expression of key pluripotency genes in the different lines and control BJ fibroblasts and H1 ESC line.

Presence of pluripotent markers still cannot assure these cells are able to differentiate into all three germ layers – the distinctive characteristic of pluripotent cells. In order to evaluate this ability, two different tests are routinely used –EBs formation and teratoma formation. They rely on similar foundations but while the former is performed *in vitro*, the latter is *in vivo* by injection of the iPSCs in mice. Spontaneous differentiation of the cells is induced, and due to their pluripotency, cells from all germ layers arise.

EBs were generated from all iPSC lines, and from analysis of the differentiated culture we could observe down regulation of pluripotent genes and upregulation of different lineage specific genes. Complementary to the *in vitro* data, teratomas from the different lines arose in the mice and consisted of tissue derivatives of the three germ layers (Figure 20A). Here we confirmed both the *in vitro* and *in vivo* pluripotency of the iPSC lines derived.

A recent substitute for these tests has recently been developed by Meissner's group¹⁵⁶ consisting of a scorecard that is able to predict the pluripotency and germ layer bias of cell lines. We decided to perform this third test and evaluate the pluripotency of our different lines. In agreement with what we saw using the traditional pluripotency tests, all lines scored as pluripotent with differentiation score averages within the ranges considered as true tri-lineage potential (Figure 21).

A



B

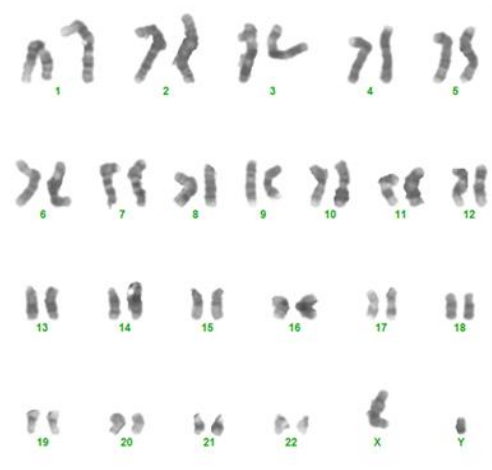


Figure 20 - Characterization of pluripotency potential of different lines Ability to differentiate towards the 3 lineages assessed by teratoma formation assay in iPSC lines from different reprogramming methods. Genomic stability assessed by karyotyping A. H&E stained slides of teratomas formed from injections of iPSCs into the testis capsule of NOD-SCID mice. Tissue derivatives indicative of the three germ lineages were observed. B. Representative image of karyotype 46, XY. BJ-pla-iPSCs at passage 15 is shown.

Further tests confirmed the normal karyotype of the different lines, as seen by the example of BJ-pla-iPSC (Figure 20B). From the analysis of the iPSC lines we can conclude that all methods are able to reprogram to a pluripotent state where the differentiation potential is kept

with a seemingly conserved genetic integrity. Regarding any method specific abnormality, in our sample population, there is no difference detected with the characterization done. Further genetic tests, such as SNP or CNV analysis, might uncover aberrations that karyotyping cannot detect. Also, a bigger cohort of samples would be necessary if method specific aberrations were found.

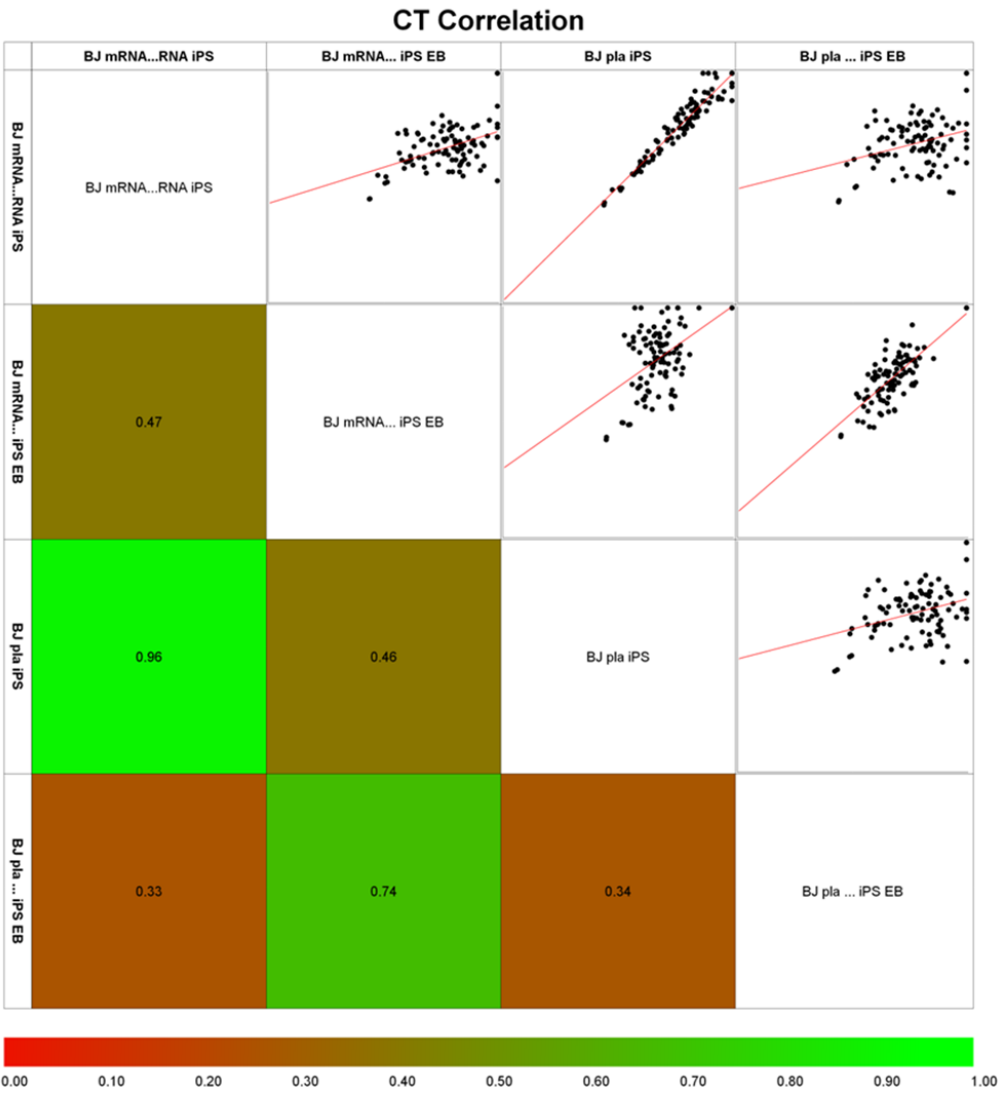


Figure 21 - Human pluripotent stem cell scorecard assay results. New assay that allows to look at pluripotent potential of iPSC and ESC lines based on a qPCR assay. Results comparing BJ-pa-iPS and BJ-mRNA-iPS lines.

3.6 Discussion

From the 3 reprogramming methodologies we chose to perform a comparative study, mRNA is the best in terms of efficiency. Nevertheless, the inefficiency at reprogramming patient lines (as observed by others^{157, 158}) renders the previous finding innocuous, since iPSC technology's greatest innovation was the ability to generate pluripotent cells from patients own cells. Another disadvantage of this method is the high cost and demanding process, requiring daily transfections for 12 days, thus making it more difficult to develop a good manufacturing practise (GMP) compliant process.

Episomal plasmids overcome the issues that mRNA presented, as it is able to reprogram patient lines. Efficiencies although not as high, are sufficient to use this method routinely to derive patient lines. Also the low cost, ease of the process and the facility to obtain GMP grade plasmids make it highly desirable.

But efficiency is not the only factor that needs to be taken into account when the ultimate goal is clinical translation. All 3 methods we compared were chosen under the premise of their safety – the most important factor – but another aspect that needs to be taken into account is the contamination from animal products. iPSCs must be derived and cultured under fully defined xeno-free reagents to comply with GMP requirements. In the next chapter we will depict the work done in deriving a method that is fully xeno-free and integration-free that can be then used for derivation of the first GMP grade iPSC line.

Chapter 4 GMP compliant reprogramming method

4.1 Introduction

Integration-free iPSC lines are an important step towards their clinical use but there are other issues that need to be addressed for generating clinical-grade stem cells such as complying with current GMPs. Two different methods can be accepted for this purpose: 1) derivation of cells under GMP requirements and 2) conversion of cells derived under research-grade condition to GMP quality standards. The first methodology grants higher safety to the cells and less work as no tests are required to determine any contamination from adventitious products – the use of GMP grade products assumes no contamination. GMP are the practices that guarantee the minimum requirements that a product (pharmaceutical or not) must meet to assure high quality and minimal risk to the consumer. It covers both manufacturing and testing of the final product.

The development of the ESC field and later the iPSC field relied greatly on the use of animal products^{1, 14, 133}. These facilitated the derivation and culture of the cells, but as the promise of using these cells as medical products came to light, the need for safer and GMP compliant methods became evident. Besides contamination with adventitious agents, the use of animal products containing undefined components presents two main problems: 1) unknown factors that can affect reprogramming, growth and differentiation of the stem cells, and 2) variability among cultures (not only among labs but inside the same lab).

With increased knowledge of stem cells, their properties and their general requirements, the field of stem cells has observed great advances in the products used for their derivation and maintenance. A race for the derivation of the first GMP grade iPSC is underway, with several findings in the past that allowed for a fully defined, xeno-free and GMP compliant method to have been derived¹⁵⁵.

4.1.1 Culture conditions

Culture conditions serve to provide the desired signals to the cells, be it for their maintenance, differentiation or even reprogramming. *In vivo*, cell fate is regulated by a combination of intrinsic and extrinsic mechanisms. Intrinsic mechanisms include specific transcription factors expressed by the cells while extrinsic mechanisms are normally related to the signals the niche provides to the cells, including contact with other cells, the ECM and growth factors. It is the extrinsic mechanism that we try to replicate *in vitro* through the culture conditions.

A great deal of work has been done on culture conditions for stem cells, with most of conditions for derivation and propagation of iPSC being based on work done for ESCs. While iPSC have only been around since 2006, the field of ESC is around since the 1980s¹⁵⁹ and has seen several advances in culture conditions. Even though no GMP grade iPSC has been derived, much of the technology for such has already been developed.

Culture media

Culture media play a crucial role in achieving a defined culture system for human pluripotent stem cells. These should contain soluble factors that act on PSCs to control their fate, both in terms of self-renewal and differentiation.

The media initially used to maintain ESC in culture contained FBS^{131, 160}, which introduces the risk of transmission of disease (e.g. prion) as well as activation of host immune system by biomolecules, such as non-human sialic acid (Neu5Gc)^{161, 162}. Xenogeneic products are also a common source of mycoplasma contamination, which are known to compromise several aspects of cell physiology and consequently affect experimental results¹⁶³. For increased safety of the resulting cells, human products were employed instead^{164, 165}. These provided an extra level of safety, as they avoid contamination by inter-species products (molecules, prions, virus, etc.) and potential immune rejection of xeno-proteins¹⁶¹. Even though they represent a step closer towards safer conditions, they still carry problems such as batch-to-

batch variability and its composition, as it is a complex mixture containing unknown compounds. Batch to batch variability is particularly problematic when it comes to albumin (one of the main components of serum), both because of the high concentration used in the culture medium compared to other proteins and its ability to bind lipids and other impurities¹⁶⁶.

The next step in improving the culture conditions of pluripotent cells was the complete removal of serum from culture. Using knockOut serum replacement (Ko-SR), several groups were able to show sustained culture of pluripotent cells^{26, 167, 168}. Serum replacement is a defined serum-free formulation optimized to grow and maintain undifferentiated pluripotent stem cells. Even though serum-free, it is known that the active ingredient in Ko-SR is lipid-rich albumin¹⁶⁹. Batch to batch variability is still a concern when using Ko-SR, and even though it allowed more standardized and better defined conditions, it still contains animal products.

Media optimization can be a daunting process but it was possible to develop a defined medium – mTeSR1 – where 18 components are added to a DMEM/F12 basal medium that was able to support pluripotent cells. This medium was first developed for feeder free ESC culture but is has since been used for iPSC¹⁷⁰. Even though the composition of the medium is known and there are no animal products, its components include human serum albumin and human sourced matrix proteins, which, as said previously, have a great batch to batch variability in their capacity to support pluripotency¹⁷¹. Despite this, mTeSR1 has been widely adopted for the culture of stem cells, being the most common one used in present days.

Further work on media optimization led Thomson's group to the development of Essential 8 (E8) medium¹⁷¹. A first attempt showed that while removal of some of the 18 factors from mTeSR (TGF β , LiCl and GABA) didn't really affect short term survival and proliferation of ESCs, removal of Albumin and other factors led to a diminished maintenance of these cells¹⁷¹. The initial study did not take into consideration the complex interactions between the different factors in the medium, as shown by the fact that Albumin is required to counter act the toxic effect of β -mercaptoethanol¹⁷¹. Re-examination of all the components and their interactions allowed the removal of some factors, while reinforcing the importance of others

(such as insulin and FGF2)¹⁷¹. The understanding of the different interaction allowed the development of E8 medium which contains 8 essential ingredients added to DMEM/F12 and is therefore completely defined. This medium supports undifferentiated proliferation of both ESC and iPSC to a comparable level to TeSR medium¹⁷¹. The development of a highly defined medium that supports stem cell culture and reprogramming facilitates the transfer from basic research of hPSCs to the clinic, with cGMP E8 medium already available commercially.

Support systems

The natural environment of stem cells niches is hard to replicate and several factors affect the success in supporting the cells in an *in vitro* environment. These include 1) presence of soluble factors, such as cytokines and growth factors, 2) cell-to cell interactions, 3) interaction between cells and the niche and even 4) physical properties of the niche, such as rigidity.

The initial culture of ESC relied on a layer of cells (feeders) to sustain stem cells. The use of animal feeder cells such as MEFs was crucial for the support and self-renewal of hPSCs by secretion of essential growth factors, cytokines and ECMs. However, the use of animal feeders carries similar problems to the use of animal sera: batch to batch variability, mycoplasma contamination, risk of transmission of disease and contamination with animal products that could lead to immunity problems if cells are used clinically. Soon after, several human cell types were tested as possible feeders and showed the ability to support stem cell growth and derivation^{164, 172-174}. Even hESC-derived fibroblasts feeders have been used, which avoids the use of other cellular sources¹⁷⁵, though not always have they proven suitable for stem cell maintenance¹⁷⁶. Despite the evolution on the feeders used for stem cell culture, the possibility of contamination from the feeder layer restricts the clinical use of ESC and iPSCs. Variability among labs and batches can also affect characteristics and differentiation ability of these cells. Furthermore, the process of using feeder layers is elaborate and costly, limiting large scale use of the cells¹⁷⁷.

And so, as an alternative to feeder cells, several ECMs or cell adhesion molecules have been assessed. Feeder-free culture initially involved the use of a complex mixture of matrix proteins derived from Engelbreth-Holm-Swarm mouse tumours – Matrigel²⁶. Matrigel provided the required support for ESC and iPSC culture²⁶. In an initial stage, Matrigel was able to support pluripotency only in conjugation with MEF conditioned medium (CM)²⁶. MEF-CM has components that were produced by the MEFS, such as growth factors, ECM and cell binding molecules that help maintain stem cells in the absence of feeders. Successful maintenance of pluripotent stem cells without the use of CM has been shown with the addition of high concentrations of FGF-2, transforming growth factor β (TGF- β 1) and Leukemia Inhibitory Factor (LIF) to the culture medium¹⁷⁸⁻¹⁸⁰. Later on, with a better understanding of stem cell culture, improved media were developed, such as mTeSR1 and E8, that if used in conjunction with Matrigel are able to maintain the pluripotent state in culture.

Even though Matrigel removes the need for a feeder cell layer and can be used with completely defined media (E8), it is still not xeno-free as it is derived from mouse tumours. Several of the components that make up Matrigel –collagen IV, laminin and fibronectin¹⁸¹ - have been evaluated in their ability to individually support pluripotency. While collagen IV and fibronectin failed to maintain pluripotency and self-renewing capability of ESC over long periods of time, laminin was successful in doing so²⁶. Laminins are proteins of the extracellular matrix, expressed early during development and are a major component of the basal lamina^{182, 183}. Several isoforms have been identified, each with its specific receptors, and that play different roles as cell adhesion proteins¹⁸³⁻¹⁸⁶. With regards to pluripotency maintenance, few isoforms have been identified with the ability to do so: laminin -111, -332 and -511/521²⁷.

The knowledge that most cell-ECM interactions rely on integrins was extrapolated to hPSCs and led to the finding that vitronectin has the ability to support hESCS self-renewal via integrin α V β 5²⁸. Additionally, when using the fully defined mTeSR1 medium, only matrigel and vitronectin were able to support ESC growth²⁸. Most of the matrix proteins capable of supporting pluripotency are expensive for large scale use, but vitronectin has the added

advantage of relative easy production and purification, making it highly appealing as a GMP compliant support system for iPSC²⁸.

The use of synthetic platforms for hPSC culture comes as an alternative to the use of biological ones, and several have been developed^{29-32, 187}. These must allow cell adhesion, spreading, self-renewal and subsequent colony formation of undifferentiated hPSCs¹⁸⁸. The combination of both strategies in a hybrid approach has also proven popular with a base polymer with biomolecules at the surface¹⁸⁷. Despite the interest surrounding these approaches there are still issues to overcome, with only a few surfaces showing the ability to sustain hPSCs for longer periods^{32, 187, 189-191}. It is important that whatever system is employed, the sterility of the system is kept. This becomes problematic when only few of these platforms show compatibility with the sterilization process, as biological motifs can be degraded or denatured by it^{187, 190, 191}.

The use of defined conditions decreases variability and so translates onto a more reproducible and reliable culture system. Several defined media and matrixes have been shown to be able to support pluripotency but the question still remains to see if they are able to support the reprogramming process.

4.1.2 Beyond GMP practises

The generation of stem cells for the use in patients is not simply a matter of getting an integration-free GMP compliant line. There are other issues that require attention if there are to be downstream applications of these cells.

These issues are varied and range from quality controls of the cells themselves to ethical and legal issues. The ethical and legal part are beyond the scope of this thesis thus will not be developed into specific details; however some of the issues surrounding the field include regulation on sourcing the donor tissue, ethical guidelines, intellectual property law and data sharing. To further complicate matters, regulations differ between national and/or regional authorities. Efforts are currently being made to form an international autonomous body with the aim of standardizing the existing regulations and proposing new regulations to accelerate the translation of PSC-based therapies to patients¹⁹².

It is also important to monitor genetic abnormalities both in iPSC and its derivatives. A lot of work has been done in addressing the genetic and epigenetic stability of these cells, with several groups showing marked genetic aberrations, while others disputing these findings⁴⁴. Initially a lot of focus was given on comparison of iPSCs with ESCs on a genetic basis. While the world was amazed at how similar iPSC were to ESC when they were first developed, from 2009 reports started recording differences between the two^{82, 84}, including at the methylation level⁸⁵. However, others studies show the difficulties to distinguish the differences between them^{90, 91, 156}.

When analysing the contradicting results, there is a tendency where the cohorts that analysed smaller sample sizes are the ones that show differences between iPSC and ESC, giving strength to the idea that the changes observed are not specific to ESC and/or iPSC but rather sample or laboratory specific. Furthermore, comparison between the original cell source and the derived iPSC have shown that most mutations are often carryover aberrations from the original cell source, mutations acquired during the reprogramming, insertional mutagenesis when using integrative methods or from the passaging process in cell culture¹⁹³. Minimizing

aberrations is important, but understating them and controlling their effects might be more important as even imperfect cells can be used⁶⁹.

When considering clinical application, immunological problems beyond the ones caused by xeno-components should be addressed, such as immune rejection. Even though iPSC technology can be applied on a personalized level, the reality of such is not easily achieved. The long and expensive process of developing a line and testing it creates obstacles on routinely using it in a patient-specific manner. Derivation of GMP-compliant iPSC lines from healthy donors that included a variety of human leukocyte antigen (HLA) variants could represent a feasible alternative, as matching of the three major types of HLA loci between recipient and donor is expected to result in less immune rejection after transplantation⁷⁸.

Much still needs to be done before stem cells are routinely used in the clinic, with only a few trials existing so far⁶⁸. The extensive work on quality controls and joined efforts for international regularization and cooperation in the field will hopefully bring us closer to clinic use of these cells.

4.2 GMP compliant method

In the previous chapter, we compared several integration-free methods of reprogramming and concluded that the episomal plasmid method was the most efficient and robust when reprogramming primary patient fibroblasts. To develop a safe, GMP-compliant reprogramming protocol, it is necessary to develop a protocol free of feeder cells and animal products that is compatible with the episomal method.

In the process of developing a GMP compliant method, we first needed to remove the use of a feeder layer and so we developed an approach where the plasmid transfected fibroblasts were transferred directly to a matrigel coated plate, in mTESR1 medium containing 0.5mM sodium butyrate¹⁹⁴. Sodium butyrate is a molecule that has been shown to increase efficiency of reprogramming. The mechanisms, by which butyrate helps increase the efficiency, promote epigenetic changes and expression of pluripotency-associated genes (methylation and demethylation)¹⁹⁵.

The above conditions gave rise to 76.33 ± 5.6 TRA-1-81 positive colonies, which could be immediately transferred and sub-cultured in matrigel/mTESR1. Under these modified conditions, the Yamanaka plasmids were even more efficient than under the feeder dependent protocol (Figure 22).

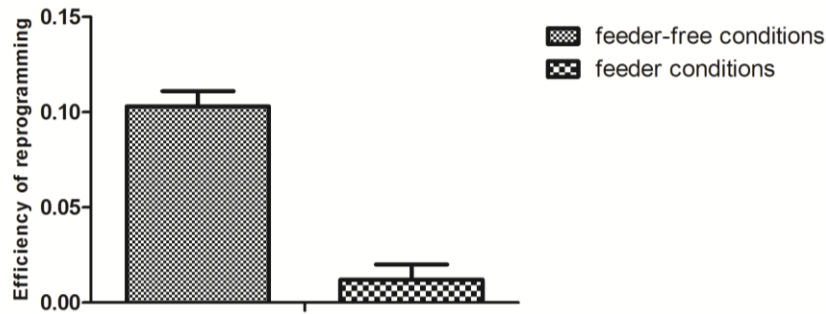


Figure 22 – Efficiency of reprogramming using episomal method in feeder Vs feeder-free conditions. Reprogramming of different human fibroblast lines under feeder (Nuff) and feeder-free (matrigel) conditions. Efficiency assessed as the number of iPSC colonies per number of transduced fibroblasts. n=3

At this stage we were able to show a method free of feeders that supports the reprogramming of human fibroblasts with our chosen integration-free method – plasmids – at efficiencies higher than when using feeders. It seems possible that the combination of Matrigel/mTeSR1 better reproduces the *in vivo* conditions of pluripotent stem cells, which translates in the higher efficiency observed.

While using Matrigel allows the removal of feeders from the protocol, it is still an undefined mixture of proteins derived from mouse tumours and thus is not xeno-free. Several ECMs including laminin and vitronectin are known to support long term pluripotent stem cell growth^{166 28 196}, however it is not known whether these ECMs can also support plasmid based reprogramming. Upon comparison of different ECMs we observed that laminin-521 was unable to support the generation of any iPSC colony. However, both vitronectin and Synthemax allowed robust generation of patient specific iPS colonies at efficiencies comparable to Matrigel (see Figure 23). Synthemax is a synthetic vitronectin-based peptide substrate for culture of hPSCs among the few compatible with the sterilization process crucial for stem cell culture¹⁹¹.

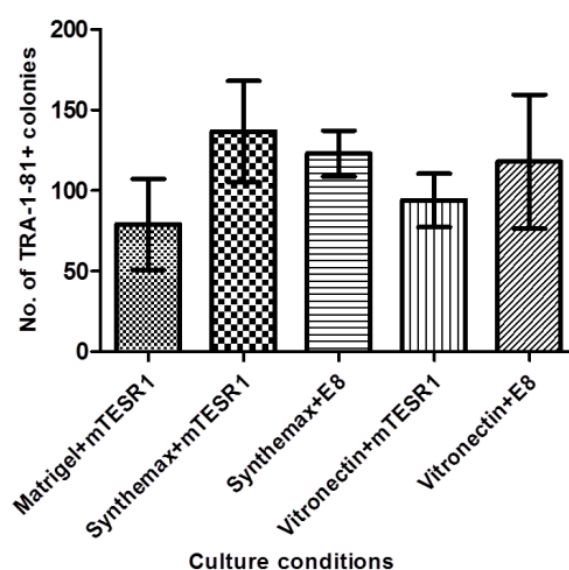


Figure 23 - Assessment of the best xeno-free matrix and medium for iPSC reprogramming. Comparison of different culture conditions in their ability to sustain iPSC reprogramming from human fibroblast. Reprogramming assessed as the number of iPSC colonies that stained positive for TRA-1-81 (crude evaluation).

Once a xeno-free support system for the derivation of iPSC was established, it was necessary to replace mTeSR1 medium for a completely defined xeno-free medium. The development of E8 medium by the Thomson group¹⁷¹ was crucial as it was the first medium to satisfy all the criteria. Again, several patient iPS lines could be derived under E8+vitronectin or E8+synthemax that could be picked and continually expanded using the EDTA method of passaging. So, replacement of mTeSR1 medium with E8 did not affect efficiency of reprogramming (see Figure 23).

Morphological analysis showed that colonies derived in E8 + vitronectin were different than those maintained in matrigel+mTeSR1, with a more flattened appearance and less defined borders (see Figure 24A). Interestingly, colonies derived in E8+synthemax showed no apparent differences (see Figure 24A). This discrepancy in morphology between iPSC derived in vitronectin and synthemax is surprising as both are vitronectin based. Others have also shown the ability to reprogram using vitronectin + E8, with colonies showing similar

morphological features¹⁶⁶. Nevertheless, it is possible that the use of a synthetic platform where vitronectin is anchored might provide better configuration for cell-ECM interactions.

As morphology is not a stringent test for pluripotency, we compared other aspects of colonies derived in mTeSR1+matrigel and E8+vitronectin. Flow cytometry analysis showed similar percentage of TRA-1-60 and TRA-1-81 surface antigens expression (see Figure 24B) while quantitative PCR showed similar expression levels of Oct4 and Nanog (see Figure 24C). For proper analysis of the pluripotent potential of the cells, teratoma and EBs formation are the most common methods used. And so, when we compared cells under the different conditions, we observed that all were truly pluripotent in their differentiation ability with cells from all 3 germ layers present. Similarly, Chen *et al.* also observe that iPSC derived with vitronectin + E8 present true pluripotency despite the colonies showing small morphological differences¹⁶⁶.

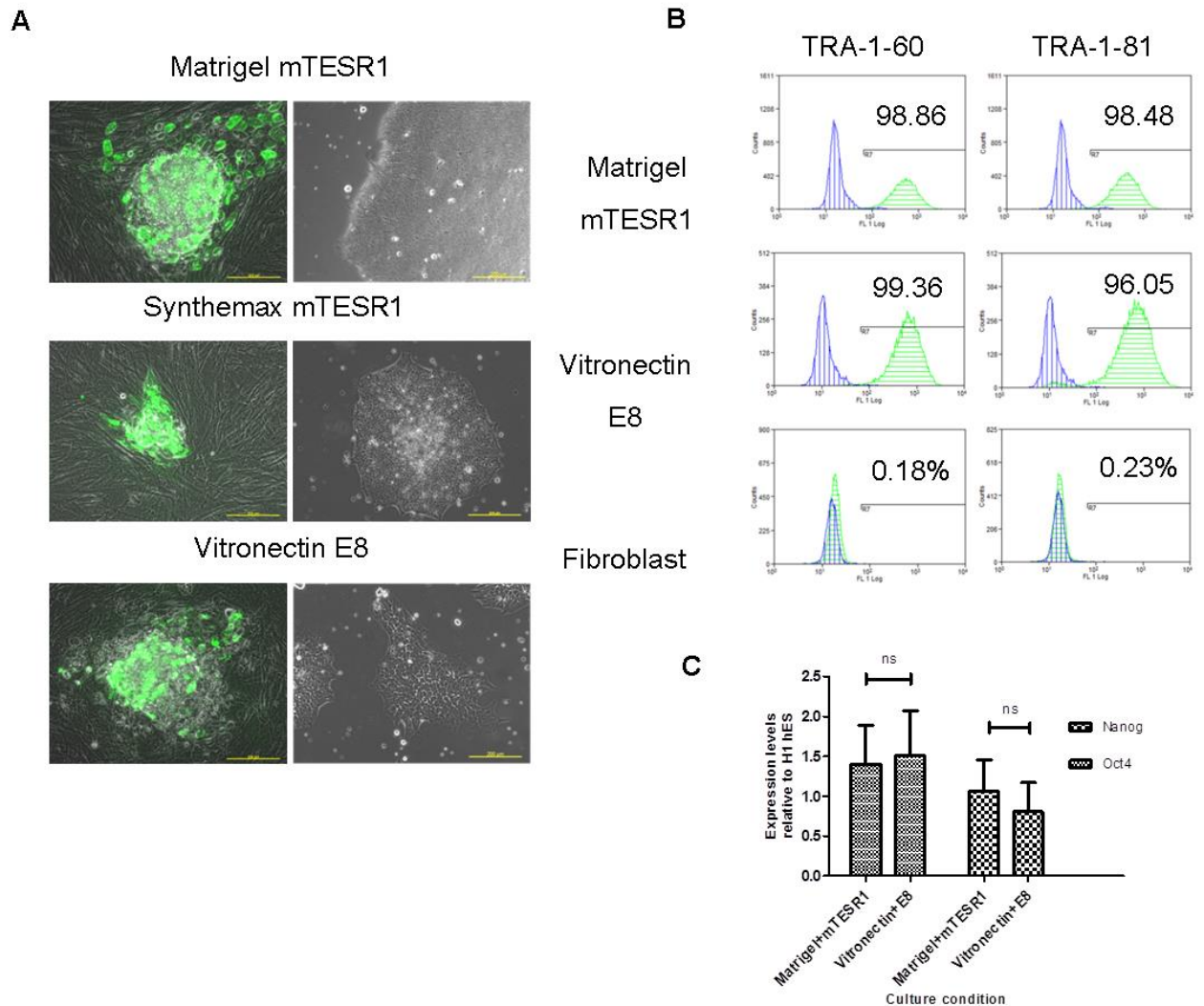


Figure 24 – Comparison of episomal reprogramming under different culture conditions A. Morphology and TRA-1-81 staining of iPSC colonies reprogrammed in the different culture conditions B. Flow cytometry analysis of TRA-1-60 and TRA-1-81 surface antigens expression of fibroblasts (negative control) and iPSC reprogrammed in matrigel+mTESR1 and vitronectin+E8 medium. C. Quantitative PCR of OCT4 and Nanog expression levels of iPSC reprogrammed in matrigel+mTESR1 and vitronectin+E8 medium.

Therefore, it seems that no obvious differences in pluripotency exist between the two conditions, making E8 + vitronectin/synthemax a desirable combination for the reprogramming process. Currently, Synthemax is the only synthetic platform commercially available and its elevated prices can hinder their use, mainly in scale up situations, since efficacy and safety are not the only considerations as costs can also affect technology adoption.

The success in reprogramming skin fibroblasts does not mean this method is compatible with other cell types (eg blood cells) since different cells can rely on different signals and/or adhesion molecules for correct reprogramming to occur. Hence, if this process is to be used with a different starting population of cells, further optimization might be required. Variability in the reprogramming process is expected when changes are introduced since different signals can be provided in a range of ways (as soluble factors, secreted by feeders, provided by the ECM, etc). This can be observed by the success of Nakawaga *et al.* in reprogramming using laminin 511¹⁹⁷. This is in contrast with what we saw, nevertheless Nakawaga was able to do it using laminin in conjugation with StemFit medium. This is of particular importance as the most likely explanation for the successful reprogramming is the fact that this medium contains albumin. Albumin has many biological and physical roles, with proven efficacy in the reprogramming process as it was frequently used in the past to aid the reprogramming process.

We have shown here a method of reprogramming that combines an integration-free process (episomal plasmids) with fully defined, xeno-free culture conditions, that is amenable to GMP process¹⁵⁵. The time will soon come when the first GMP grade iPSC is derived, a first important step towards cellular therapy using iPSC derived products. However, ultimately these hPSCs will be directed to specific cell lineages for various applications in tissue engineering and regenerative medicine. The same quality control and safety regulations will need to be applied for the direct differentiation of these cells as it was for all the other steps.

4.3 What comes next?

The development of GMP compliant iPSCs has as an ultimate goal: their use in cellular therapy. iPSCs cannot be used in their pluripotent state not only because they carry tumorigenic risk, but also because there is no cell in the adult human body that is pluripotent. There are stem cells in the adult human body such as haematopoietic stem cells but these are multipotent and not pluripotent. And so their use for cellular therapy relies on the *in vitro* differentiation towards a desired lineage.

Work has been done on differentiating PSCs to virtually every cell of the human body with different levels of success. It is not surprising that *in vitro* conditions are not enough for correct phenotypic differentiation of the cells as we know that the natural *in vivo* environment is very complex relying on intrinsic and extrinsic methods for cellular specification. Understanding of these signals has greatly improved, much due to animal work, and it is due to this knowledge that differentiation protocols were developed.

There has been a major interest of PSCs and their differentiation in the context of liver disease (with numerically the greatest number of publications) mainly in hepatocyte differentiation work. Additional advantages of iPSCs derived hepatocytes apply in the field of disease modelling and drug toxicity. In the next chapter we focus on the hepatic differentiation process and characterization of the hepatocyte-like cells.

Chapter 5 Hepatic differentiation of human iPSC

5.1 Introduction

Cellular therapy in the context of regenerative medicine is closer than ever to a routine use. The advances in the field of stem cells, mainly with the discovery of iPSC in 2006¹⁴, have allowed for autologous, patient specific therapies to become a reality. The improvement of damaged organs following stem cell transplantation as shown by several studies^{75, 198-200} gives strength to this line of research. Virtually all organ systems could benefit from cellular therapy for the restitution of normal function from damage or diseased conditions. Applications in liver disease, if therapeutically successful, would revolutionize the field of clinical hepatology.

The liver is the largest gland in the body and is responsible for both endocrine and exocrine functions²⁰¹, making it an important regulator of normal physiological processes and therefore vital to life. Due to this, liver disease is associated with high rates of morbidity and mortality which results in a high economic burden. Currently liver transplantation is the only curative procedure for patients with end stage liver disease. Unfortunately the availability of livers for transplant is not enough for the demand. Hepatocytes are the main cell type of the liver accounting for 70-90% of total liver mass, and upon transplantation they have been shown to be able to reconstitute the liver and even the correction of metabolic disorders²⁰². The large number of cells required and the limited availability of hepatocytes makes the generation of hepatocyte-like cells (HLCs) from PSCs highly appealing.

But cellular therapy is not the only application these cells have, even though it is considered the most exciting one. The generation of functional hepatocytes has interests in areas such as drug screening, disease modelling and even human bio-artificial liver (BAL) construction.

The use of hepatocytes in the pharmaceutical industry is not a new endeavor and allows expedition of novel human drug development. With iPSC technology and the ability to obtain lines with specific genetic backgrounds, it is possible to take into account variability in drug metabolism (as due to P450 polymorphisms) while testing new drugs.

As iPSC technology allows generation of lines in the context of specific genetic backgrounds, several lines were derived from patients with inherited liver diseases²⁰³⁻²⁰⁵. When these lines are taken down the hepatic differentiation, the characteristic abnormalities that recapitulate key pathological features are observed. This shows the feasibility of disease modelling in the context of the liver, which not only helps to increase the knowledge on the disease phenotype but also provides an *in vitro* system to search for therapies.

BALs work as an *ex vivo* device that takes over the liver functions, working as a bridge to transplantation or regeneration. These devices require living cells that express liver specific functions, but with it important questions arise such as what type of cells, how many, from which source and cultured in what form. Most BALs focus solely on hepatocyte population as these cells are responsible for most of the liver functions, even though other cell types (kupffer and hepatic stellate cells) have shown to contribute in cytokine and growth factor production. When it comes to the source of hepatocytes, human are the best option as non-human, even though readily available, carry risks of compatibility and contamination. Scarcity of hepatocytes and problems associated with foetal and cancer derived lines makes the possibility of stem cell derived hepatocytes for BALs highly appealing. Despite the barriers that still exist for the therapeutic use of BALs, these may be easier overcome than transplantation as the *ex vivo* setting allows an extra line of safety for the patients.

Regardless of the application, the main constraint remains the difficulty in sourcing and maintaining viable hepatocytes. Despite the regenerative potential that hepatocytes show *in vivo*²⁰⁶, such as their proliferative ability, the difficulty in maintaining and expanding those cells in culture present a major hurdle. The artificial environment created *in vitro* does not

provide the signals to maintain hepatic phenotype, and we witness a loss of their characteristic features and differentiation soon after culture²⁰⁷.

Stem cells have the potential to differentiate into hepatocytes both *in vitro*²⁰⁸ and *in vivo*²⁰⁹. Several hepatic differentiation protocols have been developed^{63, 210-212} by which the complex stimuli of *in vivo* liver development is attempted in an *in vitro* setting. As good as these multi-step protocols are, none has been able to obtain a population of true mature hepatocytes. The HLCs exhibit many morphological, phenotypical and functional characteristics of primary hepatocytes, with varying degrees between different groups^{63, 210-212}. In most cases cells present a foetal phenotype that is marked by expression of α -feto protein (AFP). Moreover HLCs perform at a reduced level when compared to freshly isolated hepatocytes, which shows they still are not completely mature and/or functional. Even though it is commonly accepted that hepatocytes derived from human iPSC resemble the foetal stage of development they have been shown to engraft, survive and perform functions analogous to adult hepatocytes when transplanted into animals.

It might be that further technological advances in the field of cell culture are required for better *in vitro* differentiation protocols, but understanding of the *in vivo* process of liver development becomes crucial information for improved hepatic differentiation protocols.

5.1.1 Liver development

Using animal models we now have an understanding of the reciprocal tissue interactions that lead to liver and hepatocyte differentiation. These studies show that much of hepatogenesis is evolutionary conserved

Early during development three germ layers are formed: endoderm, mesoderm and ectoderm. The interactions between endoderm and mesoderm lead the endoderm to take form as a primitive gut tube, further subdivided into foregut, midgut and hindgut. Liver bud formation results from foregut endoderm and its response to stimuli from the surrounding mesenchyme. Upon establishment of the liver bud, the hepatoblasts (bipotent stem cell population)

differentiate towards hepatocytes or biliary epithelium. A simplified schematic of the process can be seen in Figure 25. Understanding of the signals required at each stage are the key features required for the development of a multi-step hepatic differentiation protocol

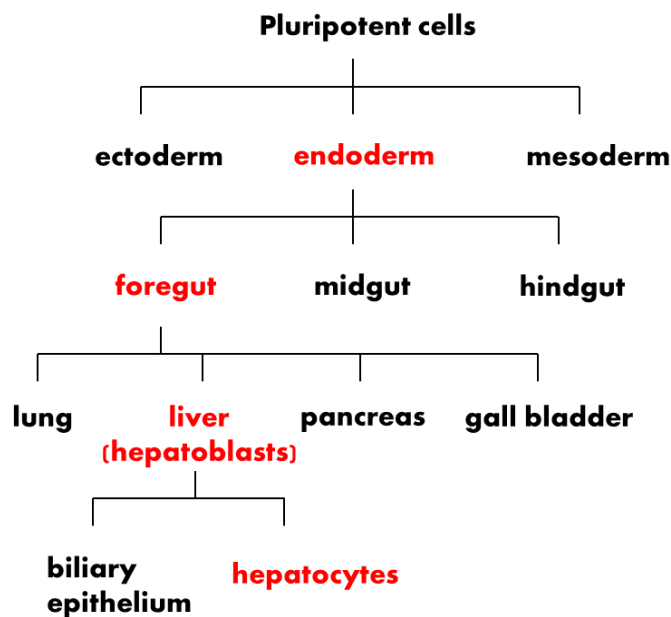


Figure 25 – Simplified schematic of route of differentiation from pluripotent state to hepatocyte

5.1.2 Definitive endoderm and foregut development

During gastrulation the 3 germ layers develop, with Nodal signalling playing a crucial role in endoderm and mesoderm determination²¹³. This happens in a concentration dependent manner with low Nodal doses inducing mesoderm and higher doses inducing endoderm^{214, 215}. Nodal signalling stimulates expression of a group of endoderm transcription factors including SOX17 and Foxa3²¹⁵.

Following gastrulation, definitive endoderm migrates to form the primitive gut tube which consists of a single sheet of cells surrounded by mesoderm. The foregut contains the common

precursors of the liver, gall bladder, pancreas and lungs²¹⁶⁻²¹⁸. Overlapping temporal and spatial gradients of Wnt, FGF, BMP and retinoic acid secreted from the adjacent mesoderm appear to regulate regional identity of the endoderm²¹⁹⁻²²³. The hepatic potential is probably due to the expression of transcription factors such as *foxa2*, *gata4/6* and *HEX*, which have important roles in early foregut organogenesis^{224, 225}.

5.1.3 Hepatic lineage

Following BMP and FGF inductive signals on the ventral foregut endoderm, liver induction occurs and the liver bud becomes apparent. Some studies suggest that BMP and FGF signalling work together in regulating hepatic specification, with analysis of downstream pathway showing that FGF regulates hepatic gene expression through MAP kinase pathway and hepatic growth through PI3 kinase pathway²²⁶.

Soon after the formation of the primary liver bud, angiogenesis and vasculogenesis proceed as the bud is permeated by angioblasts and endothelial cells mediated by the Notch signaling pathway. The bipotent hepatoblasts then differentiate into either hepatocytes or biliary epithelial cells. Hepatoblasts start expressing liver specific genes such as *AFP*, *TTR* and *Albumin*.

Bearing in mind the possibilities of HLCs for research and therapy we focused our efforts in differentiating iPSCs down the hepatic lineage. Considering the extensive literature on the subject, we optimized the hepatic differentiation process and characterized the cells in their liver specific functions as way of assessing the quality of the differentiation. Considering the risk associated with PSCs therapies, mainly the tumorigenic risk, we also performed some preliminary studies on their tumorigenic risk throughout the differentiation process.

5.2 Hepatic differentiation

Hepatic differentiation protocols try to mimic the embryonic development of the liver general in a multi-step approach. Our protocol, based on work by Chen *et al.*, follows the same strategy and comprises 3 separate stages⁶³. The first stage differentiates iPSC into definitive endoderm (DE), which represents the earliest precursors of endodermal organs (liver, pancreas, lung, gut and thyroid). In the second stage, DE are differentiated into hepatic progenitors while in the third and last stage of the protocol, the progenitors undergo a maturation step. A detailed schematic of the differentiation protocol can be seen in Figure 26.

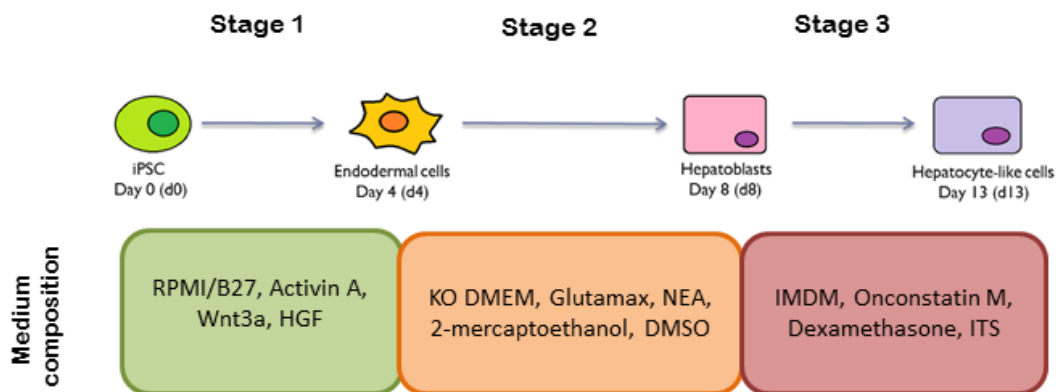


Figure 26 - Schematic of the hepatic differentiation protocol. Representation of the hepatic differentiation protocol from iPSC to HLCs. Components for each stage specific medium are provided. Briefly, cells were collected using accutase and plated at 2.6×10^5 cells/cm² in mTeSR1 medium supplemented with Y27632 ROCK inhibitor. For endodermal differentiation mTeSR1 medium was replaced with RPMI-1640 with B27, 100 ng/mL activin A, 50 ng/mL Wnt3a, and 10 ng/mL HGF for 3 days, medium changed daily. During the next step, the culture medium was replaced with hepatic commitment medium: knockout/DMEM containing 20% knockout serum replacement, 1 mM L-glutamine, 1% NEA, 0.1 mM 2-mercaptoethanol, and 1% dimethyl sulfoxide. By the end of the hepatic commitment stage, cells were collected using accutase and plated at 2.1×10^5 cells/cm², cultured in IMDM supplemented with 20 ng/mL oncostatin M, 0.5 μ M dexamethasone, and 50 mg/mL ITS.

5.2.1 Endoderm differentiation

Much work on endodermal differentiation was already available, and so we took advantage of the knowledge and combined it in order to obtain a more robust differentiation step.

In vitro endoderm differentiation requires Activin-A, a common feature to all culture systems^{63, 212, 227-230}. Activin-A activates the Nodal pathway, crucial for DE specification^{213, 214}. Besides Activin-A, composition of media for endoderm differentiation vary greatly, from the basal medium to the cytokines used. A common basal medium employed is RPMI 1640, either with or without B27, as it has shown to help endoderm differentiation as much as some cytokines. Teo *et al.* showed that using this approach they were able to obtain a population 80-95% CXCR4+ (endoderm specific marker)^{231, 232}. HGF and Wnt3a are other factors shown to help obtain a DE population and improve early hepatic lineage formation. Wnt3a together with Activin A signalling stimulates synergistic activation of Nodal and Wnt- β -catenin signalling which in turn promotes more efficient DE generation from PSCs^{233, 234}. HGF is known to have several effects on cells in culture, and for its involvement during liver development²³⁵. Its addition to Activin-A and Wnt3a further increases the synergistic effect on DE formation, improving it⁶³.

When optimizing the DE differentiation step, one possible change to Chen *et al.*'s protocol was starting from a single cell population of iPSCs. Human PSCs in general do not like to be in a single cell format and if not kept under the appropriate conditions (such as presence of ROCK inhibitor) spontaneously differentiate. The use of a single cell starting population for endoderm differentiation has also been employed by others, even adopted by StemCell technologies Definitive Endoderm differentiation kit. The resulting DE population showed equivalent endodermal marker expression to colony format. Starting from a single cell population ensures a more homogeneous final population, as it allows all cells the same conditions to differentiate. In colonies, cells in the centre and on the outside of the colony are under different pressures as they differentiate which leads to a more heterogeneous DE population (Supplementary figure 2).

We decided to compare starting in colony against single-cell format, on the quality of the DE cells. To assess the quality of the endodermal differentiation we characterized the cells in terms of endodermal specific markers. FACS analysis of cells grown as single cells and colony format shows no differences in surface marker % of SOX17 and CXCR4 (Figure 27) which shows that starting as single cells does not affect the quality of the DE cells.

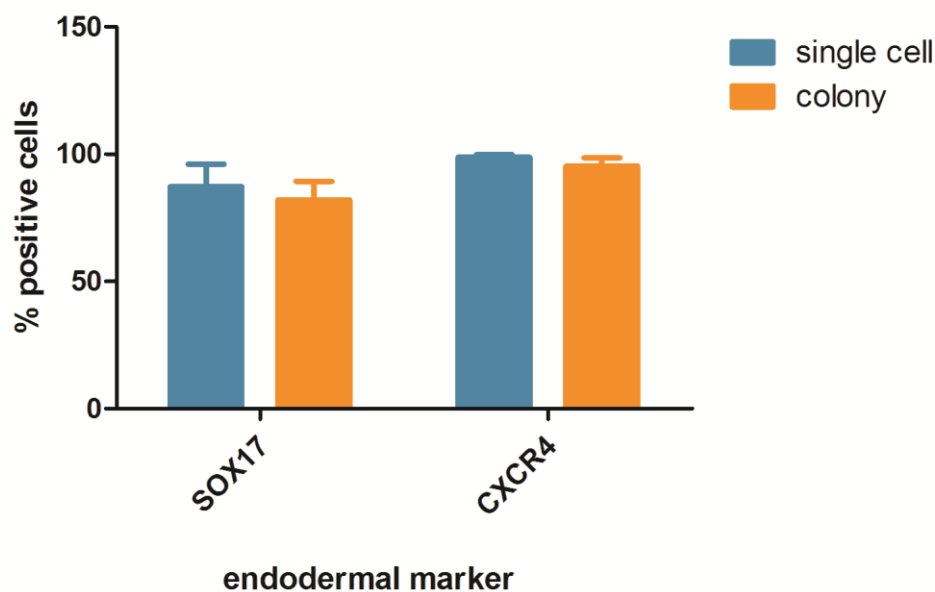


Figure 27 – Endodermal marker expression. Different iPSC lines differentiated towards endodermal cells (day 3 of hepatic differentiation protocol) and analysed for surface markers of endodermal cells. Analyses of two endodermal markers – SOX17 and CXCR4 –under different starting conditions: single-cell and colony format. Percentage of positive cells as analysed by FACS. No significant difference between single cell and colony in both markers.

Indeed by starting with a single cell population of iPSC (80-90% density), we are able to obtain a homogeneous population of endodermal cells. And so our endodermal differentiation step combines these findings, in a very efficient step with >98% CXCR4 and >80% SOX17 positive cells across different cells lines (Figure 28A, Figure 28B and Supplementary figure 3).

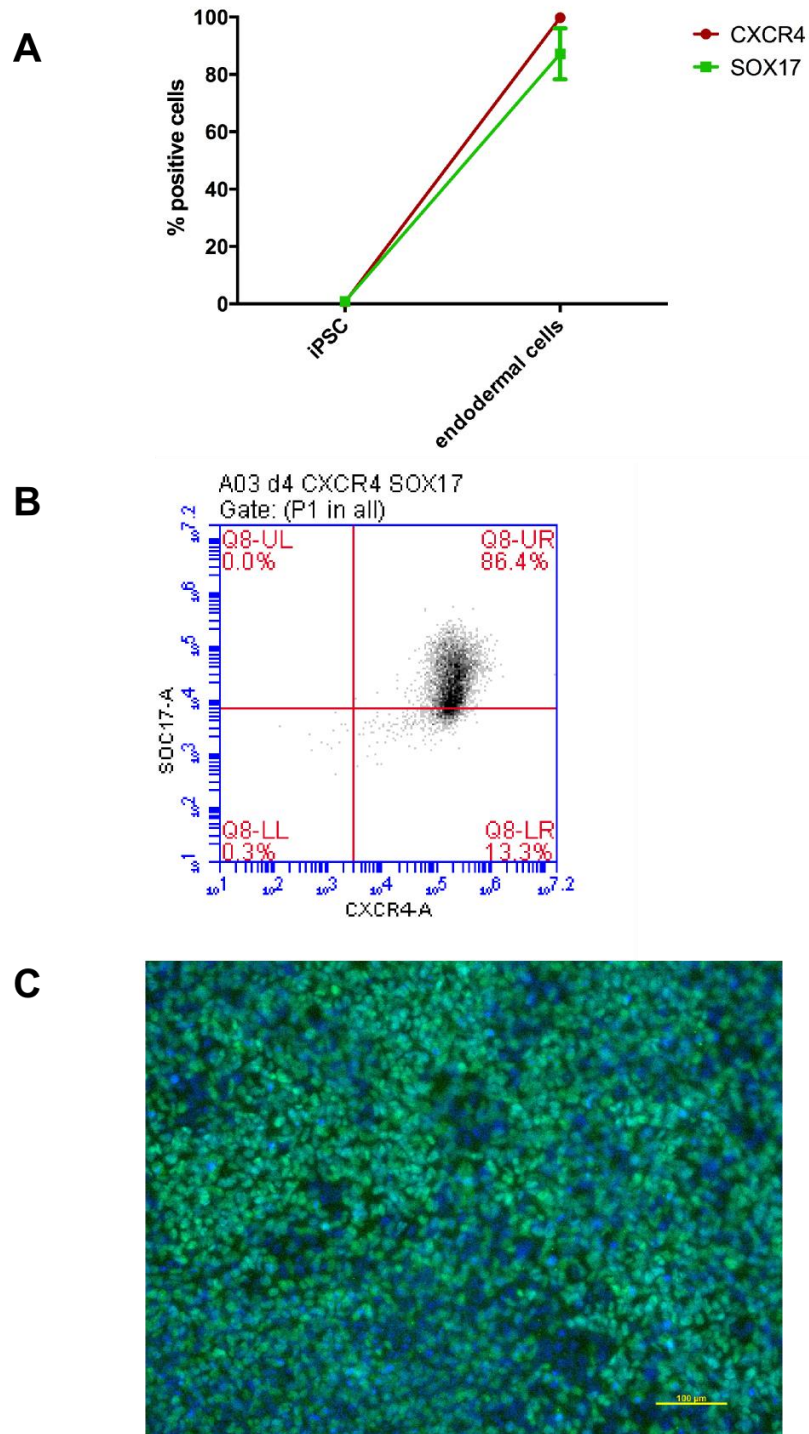
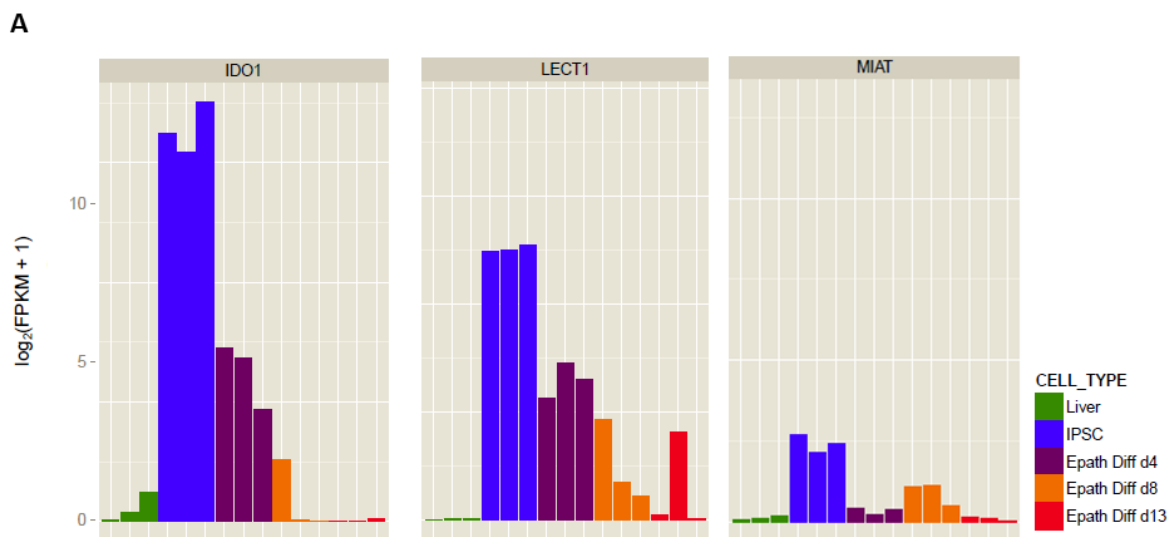


Figure 28 - Endoderm specific markers present in cells after endoderm differentiation stage. Characterization of iPSC derived endodermal cells (from single cell starting population) by detection of endoderm specific surface markers: CXCR4 and SOX17 A. Graph with average percentage of endoderm specific markers CXCR4 and SOX17 in cells from iPSC state to day 4 of the hepatic differentiation protocol (n=3) B. FACS plot showing CXCR4 and SOX17 double staining in RDP2 iPSC and cells at day 4 of the hepatic differentiation protocol. C. Immunostaining of SOX17 intracellular marker in cells at day 4 of the hepatic differentiation protocol.

Immunostaining supports FACS data, with most of the cells staining positive for SOX17 (Figure 28C and Supplementary figure 4). Gene expression analysis allows to better characterize the transcriptome of the cells by the end of the endodermal differentiation.

DeLaForest *et al.* characterized cells through the hepatic differentiation protocol and identified a list of genes specific to each of the stages of differentiation²³⁶. RNA sequencing data showed identical expression patterns to DeLaForest's, confirming the DE phenotype of cells (Figure 29 and Supplementary figure 5). Not only do we observe downregulation of pluripotent specific genes but an overall upregulation of endoderm specific genes (Figure 29 and Supplementary figure 6). Pluripotent genes show a downregulation when compared to iPSC, but some can still be detected in DE cells. This is not surprising as these cells correspond to early stages of development, and so still possess an immature phenotype. Gene expression by qPCR was also performed for a subset of genes, and overall confirms the data seen by RNA sequencing with detection of endodermal specific markers and absence of other lineage markers (Supplementary figure 6).



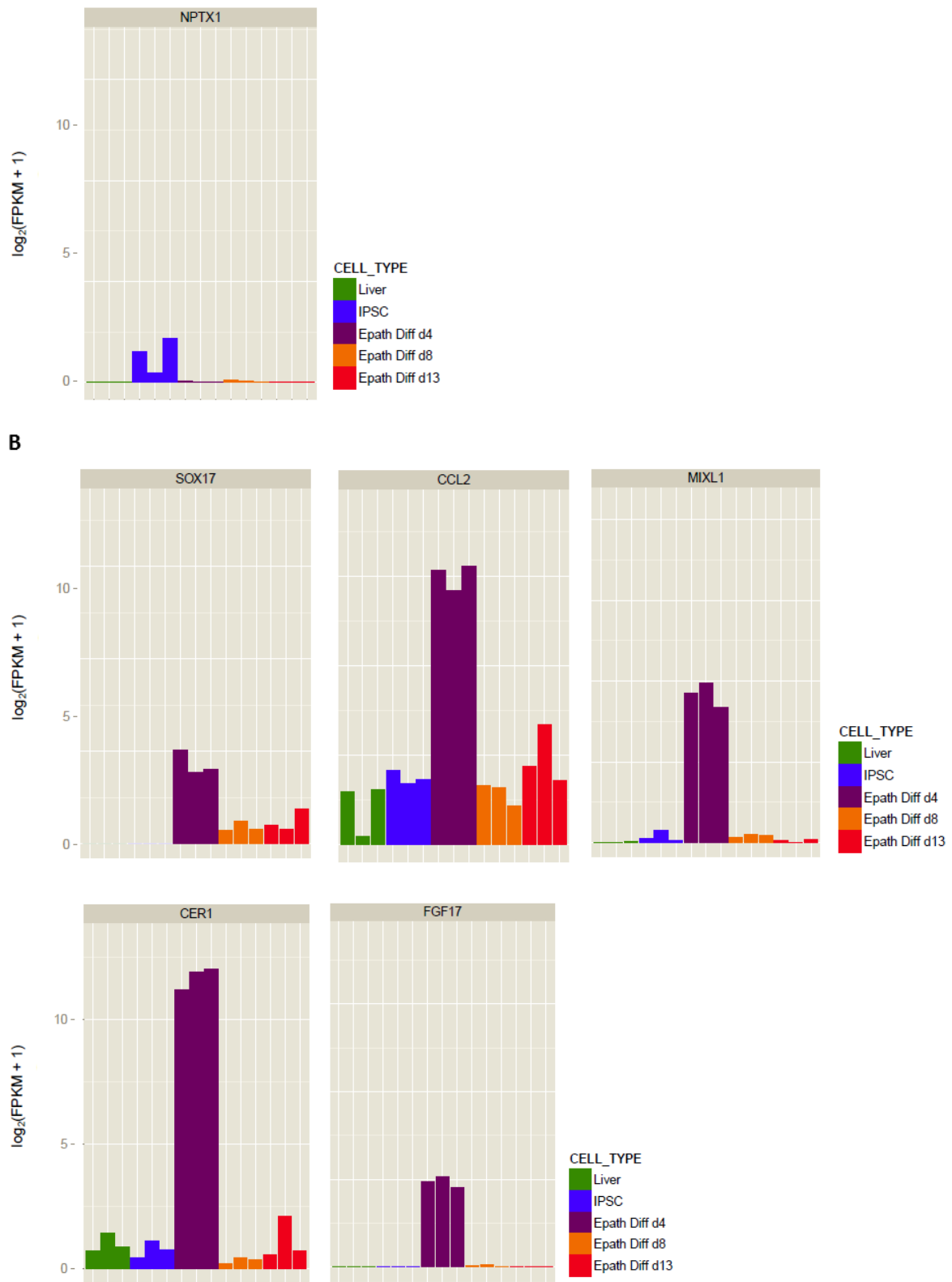
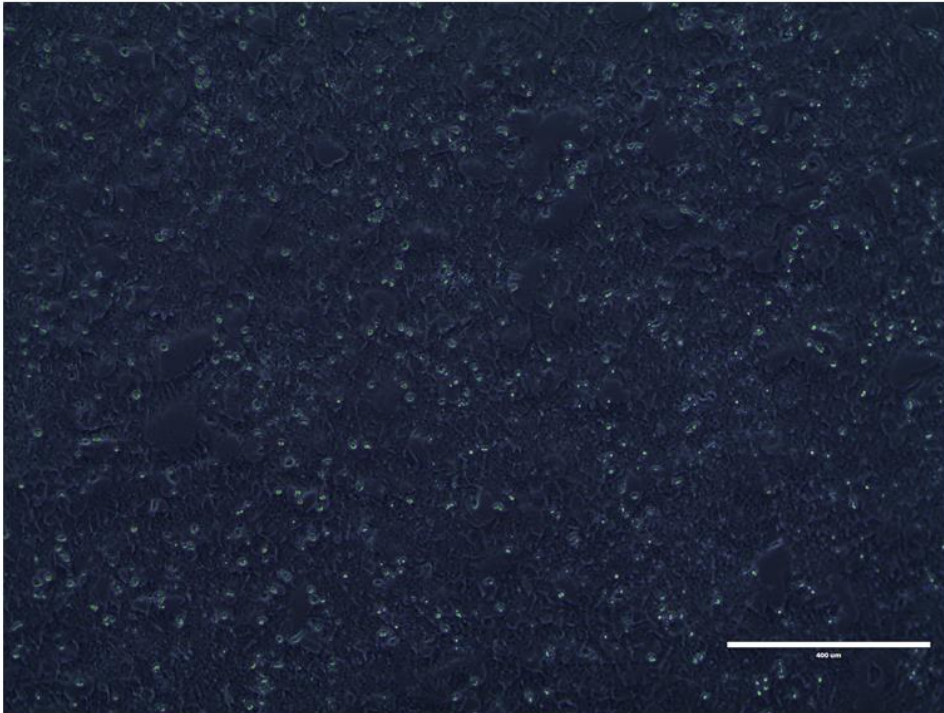


Figure 29 - Gene expression during hepatic differentiation. Gene expression of cells at different stages of hepatic differentiation and liver sample (positive control) Gene expression obtained from RNA sequencing data. Expression assessed as $\log_2(\text{FPKM}+1)$ A. Pluripotent specific genes. B. Endoderm specific genes.

During this stage we observe a high level of cell death, shown to be associated with exposure to Activin-A and Wnt3a²²⁸. Proliferation of cells is also observed, counteracting the numbers lost to cell death. The balance between these two affects density of cells in culture which, depending on the cell line, can vary (Figure 30). This variability is not surprising as iPSC, even though pluripotent, have shown to have propensity for certain lineages when left to spontaneously differentiate²³⁷⁻²³⁹.

The rapid and efficient generation of endodermal cells during this stage is important for the following hepatic commitment stage as only DE has the ability to form the liver

BJ iPSC line



RDP2 iPSC line

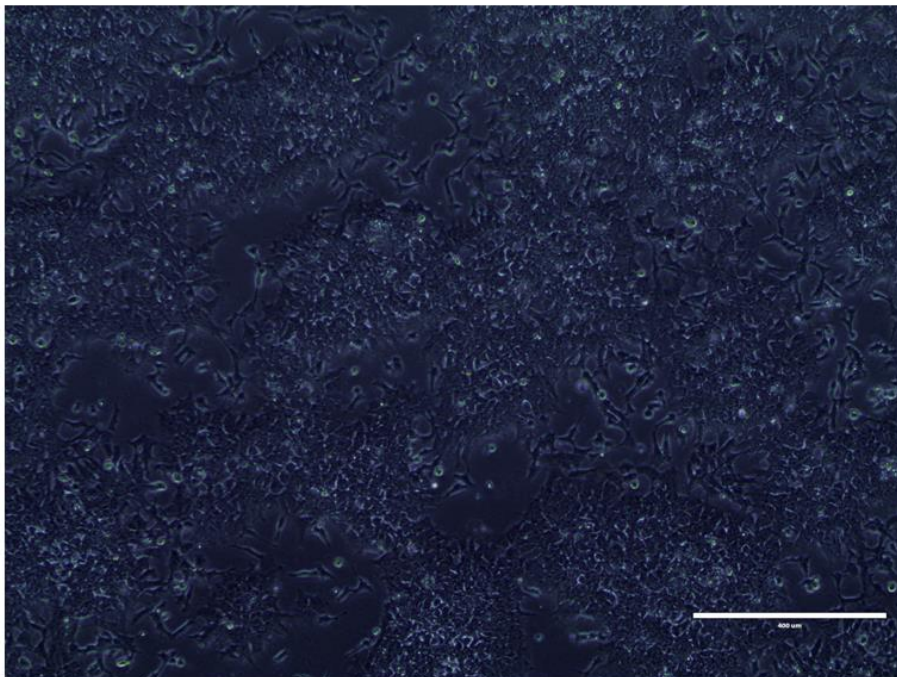


Figure 30 – Cell density at day 4 of the differentiation protocol from two different iPSC cell lines. Representative phase images of endodermal cells (day 4 of hepatic differentiation) from different cell lines. Cell densities can vary between different lines

5.2.2 Hepatic commitment and maturation

In the first stage of our hepatic differentiation protocol, human iPSC were converted to endoderm cells that are propitious for hepatic commitment. Step 2 and 3 focus on hepatocyte specification.

Following Chen *et al.*'s protocol, by the end of the endodermal differentiation stage, cells maintain a high confluency (variable between cell lines and passages) ranging from 70-100%. As cells enter the hepatic commitment stage, levels of proliferation and cell death attenuate, but cell division is still observed and cells quickly become over-confluent. Contact inhibition or lack of contact signalling can affect the cells' phenotype. Hepatocytes are particular when it comes to the seeding density, as over-seeding hepatocytes can lead to cell death, while under-seeding provides inferior functional results. As cells were too confluent, we decided to add a passaging step. The addition of this step improved the quality of the final population by giving cells improved contact signalling for the remainder of the differentiation. An added advantage of this passaging step is the possible positive effect it can have in removing PSCs as these do not enjoy being passaged and kept in a single cell state.

To assess the quality of HLCs, characterizations of their hepatic phenotype is required. Detection of both AFP and Albumin expression in cells of our differentiation protocol (Figure 31) are encouraging signs of correct hepatic lineage commitment, as these are liver specific factors. AFP is a liver specific marker expressed during foetal development, being lost after birth²⁴⁰. On the other hand, albumin is the most abundant protein synthesized by adult hepatocytes, expressed at lower levels in early foetal liver that keep increasing until these mature²⁴¹. While cells at day 8 show some level of expression of AFP (Figure 31A and Supplementary figure 4), it is at day 13 that cells present a strong hepatocyte-like phenotype with high levels of AFP, and detection of Albumin (Figure 31). The expression levels of these genes indicates a foetal phenotype as AFP is strongly present, both at mRNA and protein level, while albumin is detected at reduced levels (Figure 31). This is also in accordance to what is seen in the literature, as no complete mature phenotype has been achieved in hepatic differentiation.

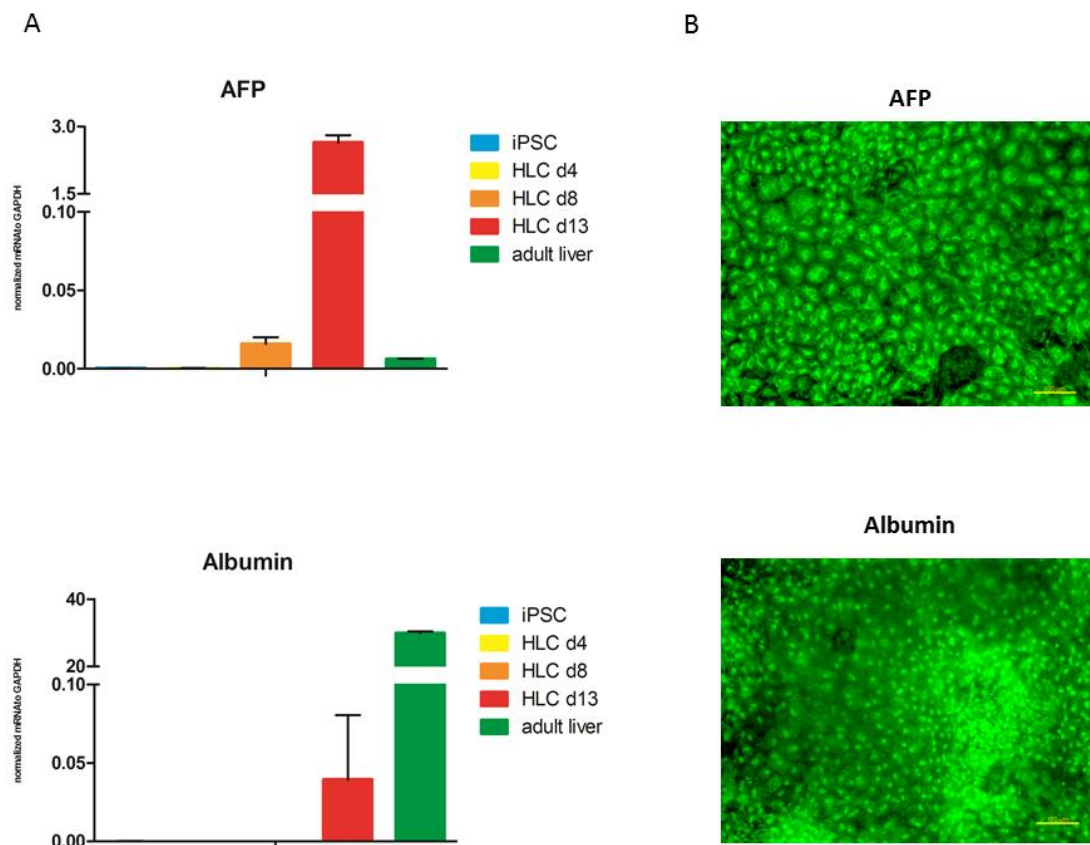


Figure 31 - Expression of liver specific markers during hepatic differentiation protocol. Characterization of iPSC derived hepatocyte-like cells by detection of liver specific markers: AFP and Albumin. A. AFP and Albumin gene expression levels during hepatic differentiation (iPSC, day 4, day 8 and day 13) and adult liver (as control) detected by qPCR. Expression levels normalized to GAPDH endogenous control. B. AFP and Albumin immunostaining of HLCs at day13 of hepatic differentiation protocol.

Besides AFP and Albumin expression, we used RNA sequencing data to compare our cells to the stage specific markers DeLaForest defined in his differentiation protocol²³⁶. RNA sequencing data again showed that our cells presented a correct expression profile for each of the stages (Figure 32 and Supplementary figure 5). Gene expression by qPCR also showed upregulation of liver specific genes such as HNF4a and TTR (Supplementary figure 6). Detection of stage specific genes in our differentiation protocol supports the claim that indeed our cells are following the hepatic lineage.

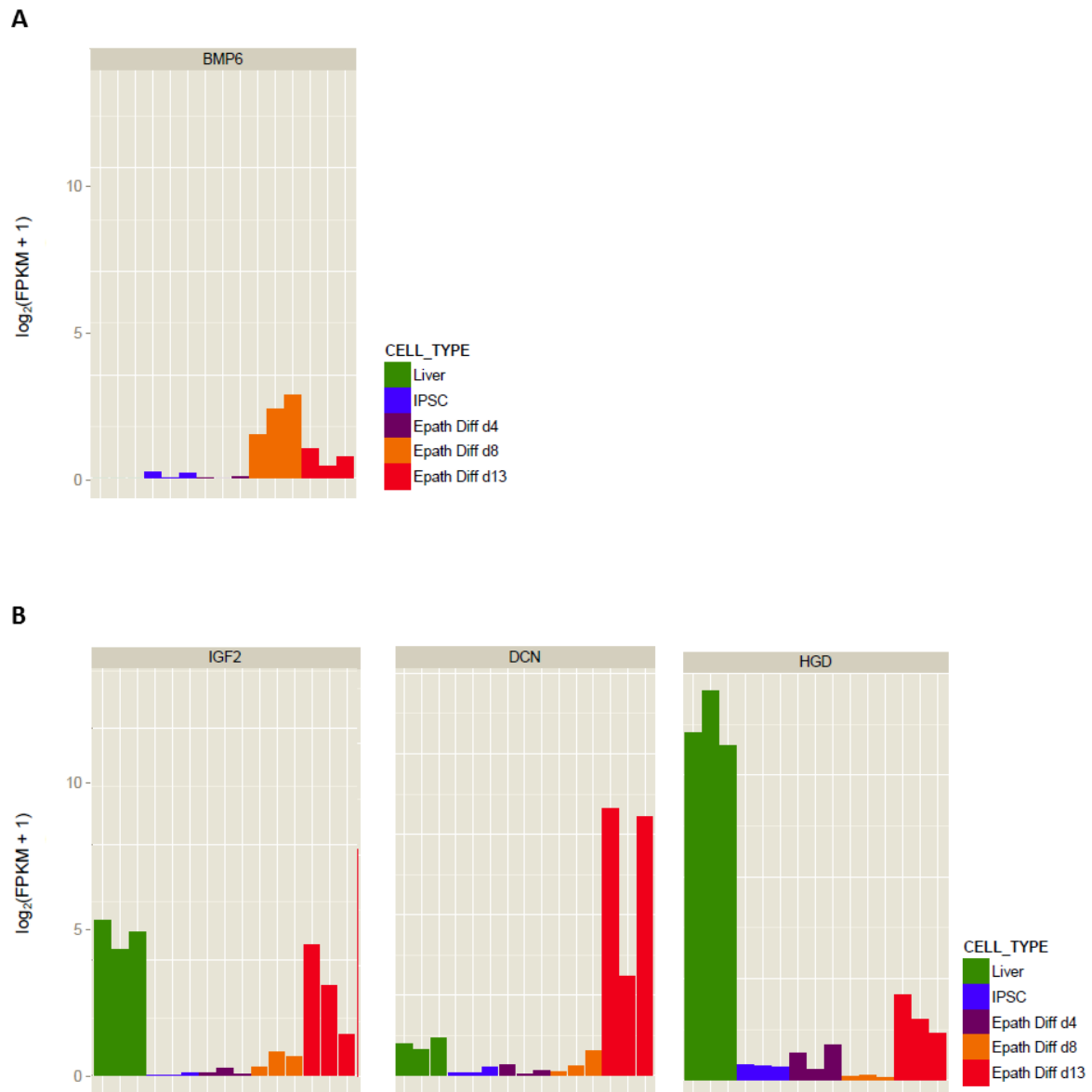


Figure 32 - Gene expression during hepatic differentiation protocol: stage 2 and 3. Gene expression of cells at different stages of hepatic differentiation and liver sample (positive control) Gene expression obtained from RNA sequencing data. Expression assessed as $\log_2(\text{FPKM}+1)$ A. hepatoblast specific genes. B. Genes turned on at the Hepatic maturation stage

Even if HLCs show a transcriptional landscape similar to foetal phenotypes by the end of the differentiation protocol (day 13), it is important to confirm that cells possess functional characteristics of hepatocytes as these are the relevant features for the use of HLCs in research and therapy. Glycogen storage is an important function of hepatocytes and can be assessed by using periodic acid-Schiff (PAS) staining. This staining method relies on

oxidation of groups in the sugars (glycogen in this case) that when react to the Schiff reagent give a purple/magenta colour. HLCs at day 13 show ability to store glycogen, shown by the presence of purple/magenta stained cells as seen in Figure 33A. PAS staining also shows us that the HLCs consist of a heterogeneous population as not all cells are able to store glycogen.

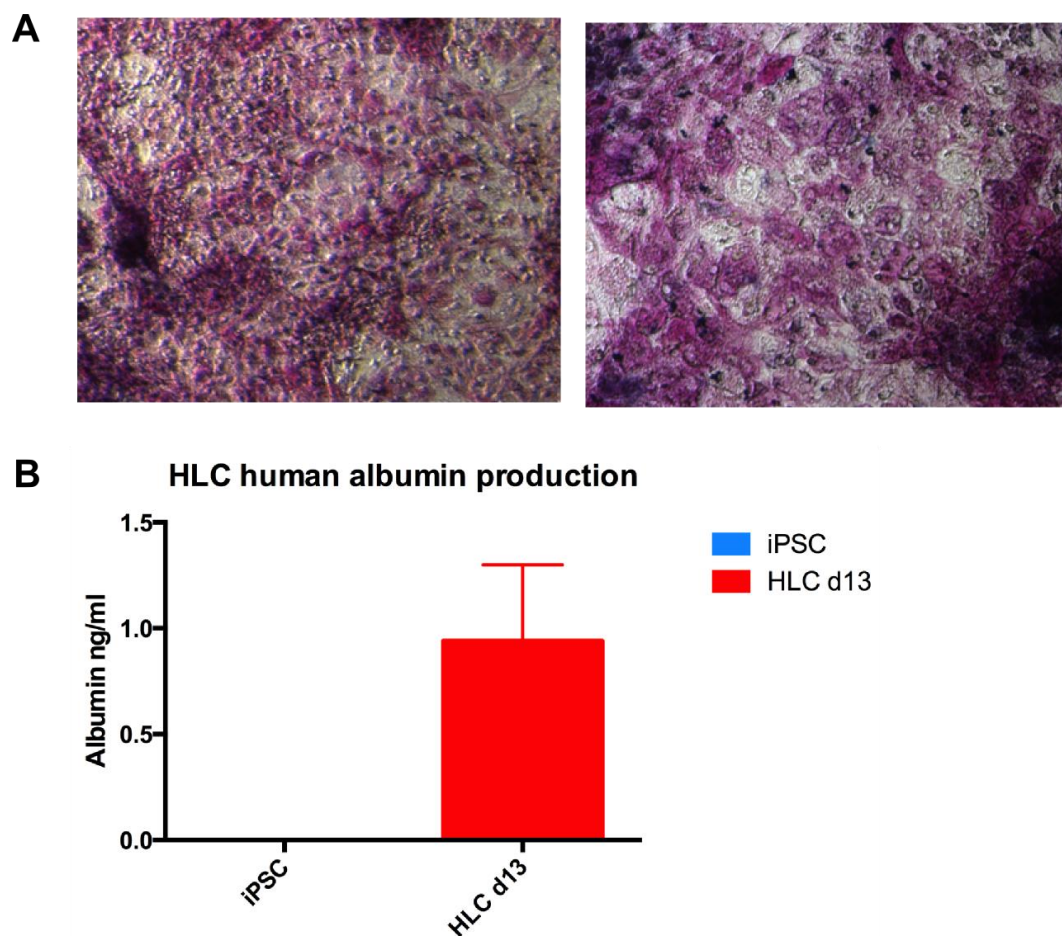


Figure 33 – Functional characterization of HLCs. Characterization of HLCs based on hepatocyte known functions such as albumin production and glycogen storage A. PAS staining of HLC d13 as a functional test of HLCs ability to store glycogen. B. Albumin production by HLCs d13 as assessed by human Albumin ELISA Kit. iPSC as negative control for Albumin production.

Albumin production is another functional test usually used to assess quality of hepatocytes in culture. Albumin is the most abundant protein synthesized by adult hepatocytes, and despite previously showing expression and transduction of this protein in HLCs day 13 (Figure 31), we checked albumin levels present in the supernatant of cells by ELISA. As shown in Figure 33B, albumin can be detected in the supernatant of cells even if at low levels.

Even with the presence of functional features of hepatocytes, HLCs present a foetal phenotype and consist of a heterogeneous population, features that can affect the transfer of HLCs for clinical use. This is common to several hepatocyte differentiation protocols, even if improvements have been made. This can be a problem of the differentiation protocols but might also result from the stability of the hepatic phenotype in culture. It is known that long term culture affects quality of primary hepatocytes²⁴², and the same mechanisms that affects primary hepatocytes in culture might affect cells of the differentiation protocol as these reach a hepatic phenotype.

Regarding the heterogeneity observed, it is possible that cells find themselves at different developmental stages during the differentiation, with some reaching a hepatocyte-like state earlier than others. Another possibility is that the heterogeneity is actually a requirement for correct hepatic differentiation, as different cells are necessary for production of the proper signals required for hepatic differentiation.

Quality of HLCs has routinely been assessed by comparison to reference controls such as cell lines of immortalized hepatocytes or primary hepatocytes. Immortalized hepatocytes offer a reproducible reference but they show reduced quality in hepatocyte physiological functions²⁴³. On the other hand primary hepatocytes, even though the gold standard, show high level of variability among different samples, making it challenging to compare results across platforms. This together with the different criteria used to characterize the cells, makes it challenging to assess the quality of HLCs and the different differentiation protocols.

The *in vitro* systems might never be able to replicate *in vivo* conditions and so proper maturation of cells might require an *in vivo* step. Moreover, the best way to check HLCs for hepatocyte functions is to use them in an *in vivo* model of liver failure and check if the liver can be rescued.

5.2.3 Liver failure rescue

For a deeper functional analysis and to assess the therapeutic use of the HLCs we used them in a mouse model of liver failure.

Carbon tetrachloride (CCl₄) is one of the most potent hepatotoxins and is commonly used for both chronic and acute models of liver failure, with varied methods being employed^{63, 230, 244-247}. An optimization step resulted in liver failure by injection of CCl₄ diluted 1/10 in oil, in nude mice via inter peritoneal (IP) injection at 2.5ml/kg concentration. Upon IP injections of CCl₄, the mice showed clear signs of pain and over the 48h following hours died from liver failure. Analysis of the livers showed the devastating effect of CCl₄, with high level of cell death with disruption of the normal liver architecture (see Figure 34 A).

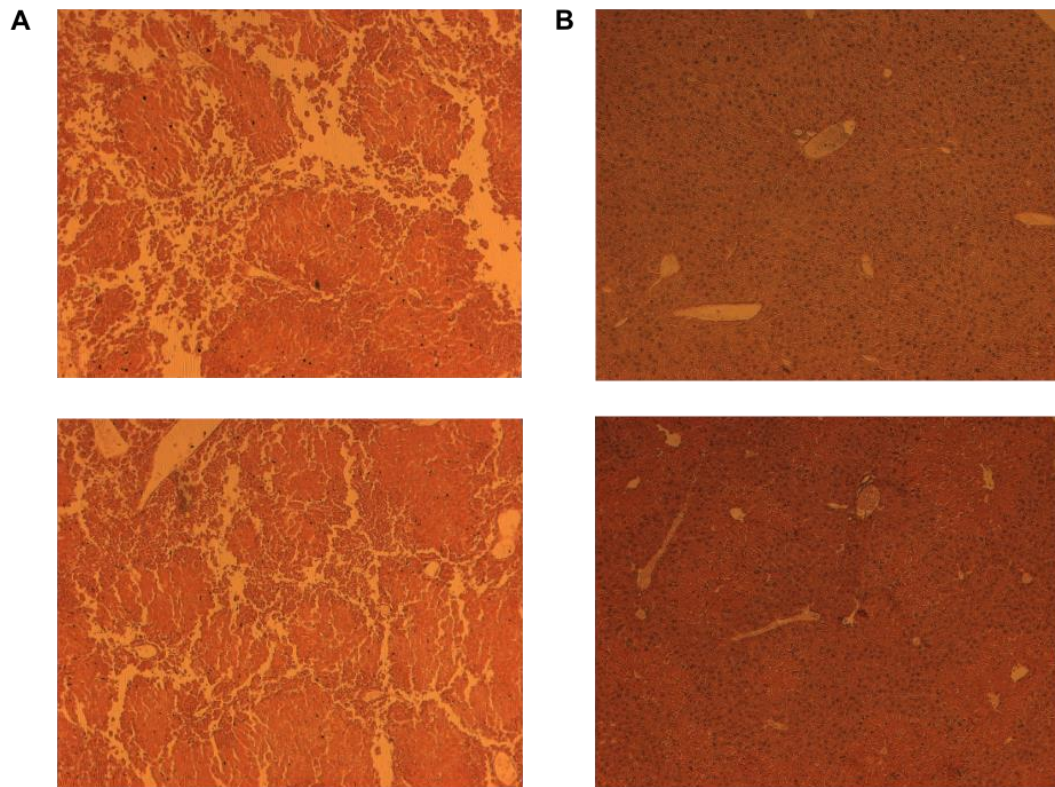


Figure 34 - H&E of mouse Livers. Fulminant hepatic failure was induced by IP injection of CCl₄ 2.5 mL/kg body weight; 1:10 v/v in mineral oil. For liver rescue, 24 hours after CCl₄ administration, mice were injected with HLCs. 1×10^6 cells/mouse diluted in PBS were administered via tail vein injection in a total volume of 100 μ l. H&E to assess damage and rescue of mice livers. A. Liver sections of mice treated with CCl₄ 24-48h after injection B. Liver sections of surviving mice treated with CCL4 and HLC transplant, 60 days post injection.

To try and rescue the phenotype and assess the HLCs, intra venous (IV) injections of the cells were performed in mice previously treated with CCl₄. This resulted in survival of 3 in 4 mice (see Figure 35A) for a period of >60 days with no signs of health problems, rescue of normal liver architecture (Figure 34B) as well as detection of human albumin in their blood at 9 weeks post HLC injection (Figure 35B). Survival of HLC injected mice and human albumin detection indicates that HLCs were able to engraft and assume hepatocyte specific functions.

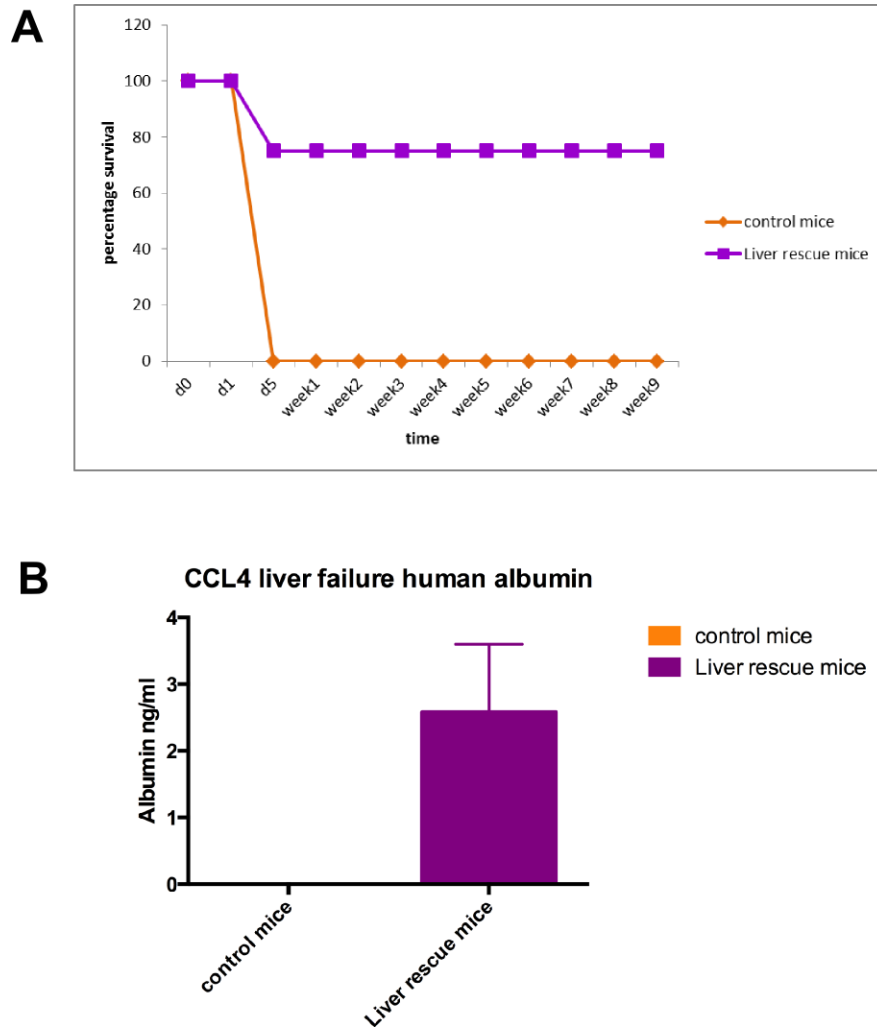


Figure 35 – HLC engraftment experiment analyses. Fulminant hepatic failure was induced and for liver rescue, 24 hours after CCL4 administration, mice were injected with HLCs. Survival and presence of human albumin in treated mice as proof for HLC engraftment A. Survival curve of mice treated with CCL4 (n=4) and CCL4+HLC (n=4) B. Albumin detection in surviving mice serum (n=3) against control mice (n=3) using human Albumin kit.

It is possible that engraftment of the HLCs and their *in vivo* survival allows for further maturation and establishment of a better hepatic phenotype than when cells are cultured *in vitro*. With this in mind, the previous experiment was planned so mice were injected with cells at different stages of hepatic maturation: HLC at day 8 and HLC at day 13 (two mice each).

Cells at day 8 are more immature, which carries a higher risk of tumour formation, but since they have already committed into the hepatic lineage they might be able to engraft, mature and take up liver functions and in doing so rescuing the liver failure phenotype. The mice engrafted with these cells showed two different outcomes: one survived and one died. The fact that one of the mice survived indicates that HLCs at day 8 are capable of liver failure rescue, although a ½ success rate is not encouraging. Unfortunately the surviving mouse presented a tumour in the spleen, which puts into question the safety of these cells (Figure 36). On the other hand, HLCs at day 13 were able to rescue the liver failure in both mice with no signs of tumours present.

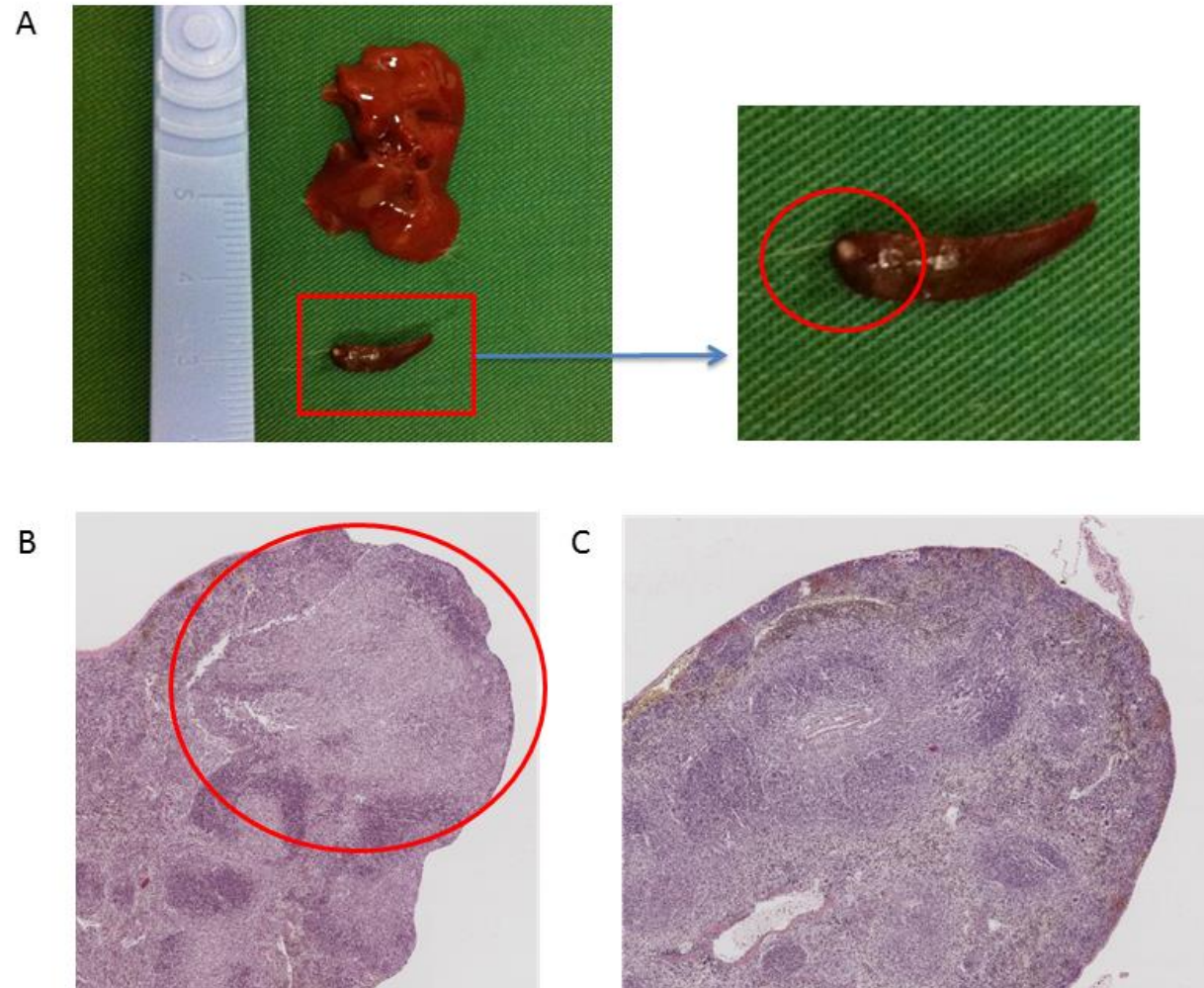


Figure 36 - Tumour in liver failure rescue with HLC. Fulminant hepatic failure was induced and for liver rescue, 24 hours after CCl₄ administration, mice were injected with HLCs. Survival and presence of human albumin in treated mice as proof for HLC engraftment A. Survival curve of mice treated with CCl₄ (n=4) and CCl₄+HLC (n=4) B. Albumin detection in surviving mice serum (n=3) against control mice (n=3) using human Albumin kit.

The low number of animals in the initial study makes it statistically weak and hard to take meaningful conclusions, mainly when it comes to the condition with HLC at day 8 as survival of ½ animals is not encouraging. And so a second study was planned with a bigger cohort of animals, following the schematic of Figure 37.

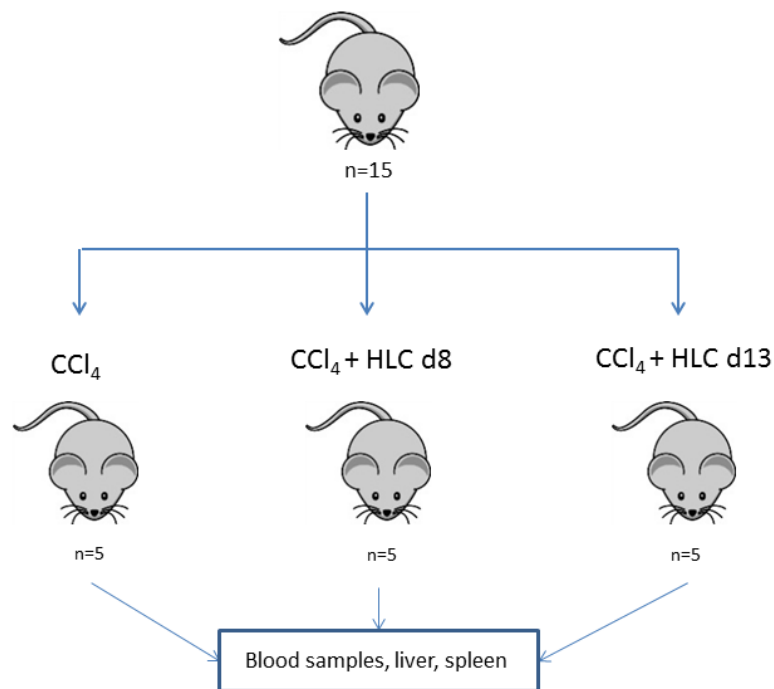


Figure 37 – Schematic of mouse liver failure experiment. After preliminary study with 4 mice, bigger experiment was planned with 5 mice for each condition: CCl₄ liver failure, CCl₄ liver failure + HLC d8 rescue and CCl₄ liver failure + HLC d13 rescue.

As previously, upon injection of CCl₄ the mice showed signs of distress but unlike before they did not die over the next 48h. By day 4-5 of the experiment we realized that the liver failure model was not successful as the control animals were still alive and showed no signs of health problems. Indeed histological analysis of the livers of these mice showed normal architecture and no sign of cell death (Supplementary figure 7). This shows that CCL4 was

not able to induce sufficient levels of liver damage and so it was not possible to assess the ability of HLCs in liver failure rescue.

The use of CCl₄ for liver failure has been extensively used in both chronic and acute models, but a big variability on the way it is employed is observed, from different concentrations, ways and number of administrations^{230, 244-246, 248}. This together with the liver regenerative capacity makes the model of liver failure with CCl₄ quite unreliable.

Lack of time and resources does not allow for further optimization of the CCl₄ model. The use of fumarylacetoacetate hydrolase (FAH) deficient mice would prove a better option for future liver failure studies. These transgenic mice are a genetic model of hereditary tyrosinemia, and develop liver failure when NTBC drug treatment is removed. This model is more reliable, as liver failure is always observed in the absence of NTBC. This would allow better assessment of liver rescue using HLCs.

Overall we have a model of hepatic differentiation where HLCs show characteristics of hepatocytes both at expression and protein level. Functional tests also show characteristics of hepatocytes such as glycogen storage. Liver rescue might be the best way to assess if indeed these cells can perform hepatocytes functions, and even though we have a reduced number of animals, we do show rescue from liver failure.

All these analyses have been routinely used to characterize the quality and resemblance of HLCs to mature hepatocytes. We thought to characterize the cells on other aspects, focusing more on their pluripotent origin than the hepatic phenotype and so decided to look at their tumorigenicity.

5.3 Tumorigenicity of HLCs

Hepatic differentiation from pluripotent stem cells carries the risk of undifferentiated or partial differentiated cells being present and contaminating the final population. Pluripotency is an important feature of iPSC but it becomes a risk when we talk about cellular therapy.

The ability to reprogram somatic cells to a pluripotent state has challenged the concept of terminal differentiation. While stability of a cell's state is crucial for survival of the organism, there is evidence showing that in a stress situation, stemness has been activated in differentiated cells. This poses the question whether the term stable differentiation isn't more appropriate than terminal differentiation. Natural occurring "stemness" activation has only been observed in lower organisms (eg: planarians)²⁴⁹ but the potential might still be there in higher organisms, we just need to tap into the right signals to do so. To tap into the *in vivo* plasticity of differentiated cells one would need to manipulate cells, their environment or both.

Another example of how the somatic state of a cell has been shown to change is in Cancer. Cancer is a very complex disease with a high degree of diversity both between cancer types and within an individual tumour. It is well accepted that cancer results from a stepwise accumulation of mutations that lead to a disruption of normal balance between cell division and cell loss. As mutations accumulate normal tissue architecture is lost with a growing mass of tissue emerging, termed tumour or neoplasm. But tumours are more than just a mass of cells, being complex structures with several cell types and its own microenvironment (Figure 38).

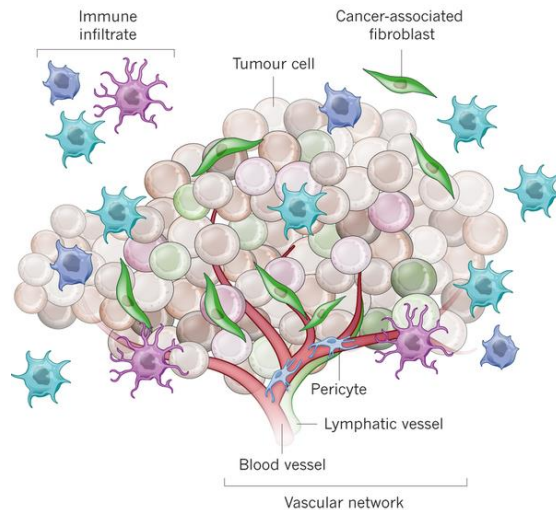


Figure 38 – Complexity of tumours. Tumour niche is dynamic, composed of several components including neoplastic cells, extracellular matrix, vascular endothelial, stromal and immune cells.

The evolution from normal cells to a neoplastic state is characterized by the acquisition of distinct and complementary capabilities that enable tumour growth and metastatic dissemination, these are termed as the hallmarks of cancer²⁵⁰. “These functions are acquired in different tumour types via distinct mechanisms and at various times during the course of multistep tumorigenesis”⁹⁹. In 2000 a paper defined 6 hallmarks for the development of cancer²⁵⁰, with subsequent research increasing the understanding of cancer and in 2011 this list was updated⁹⁹. Figure 39 shows a schematic with all the hallmarks and this can be reviewed in Hanahan and Weinber⁹⁹.

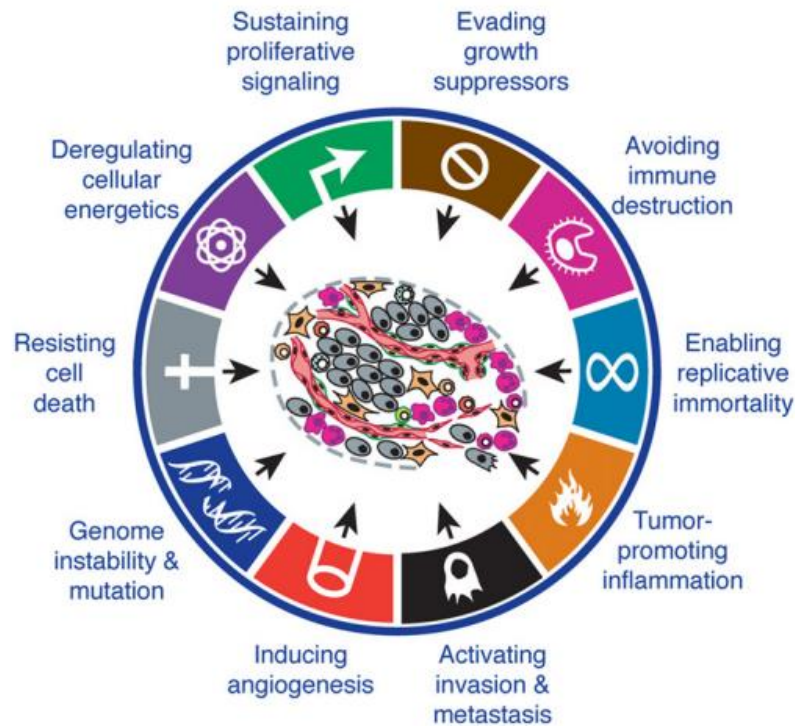


Figure 39 – Hallmarks of cancer. Adapted from original paper “hallmarks of cancer: next generation”⁹⁹

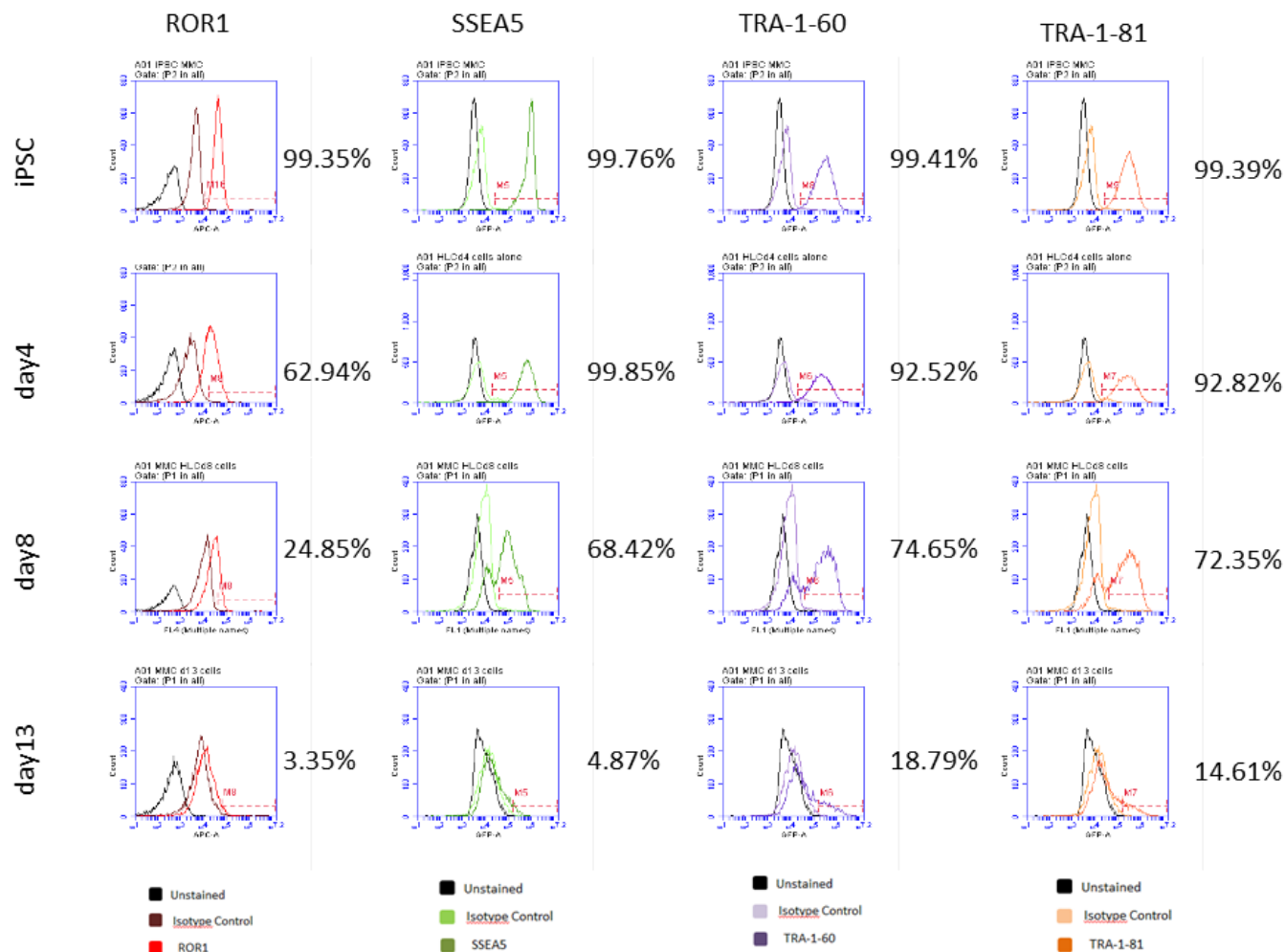
The different processes that lead to tumour formation are complex and distinct with the ultimate aim of survival and proliferation of cancer cells, which is achieved by steps such as genomic instability, avoidance of immune destruction, replicative immortality and even resisting cell death (Figure 39). The increased complexity is also marked by the presence of various cell types that work both individually and collective for tumour survival and expansion. One cellular type believed to be present in many if not most tumours are cancer stem cells (CSCs). CSCs are the cells able to form new tumours upon transplantation into a host^{251, 252}. These were initially detected in hematopoietic malignancies^{253, 254} and are usually fractioned according to cell-surface markers. Identification of a tumour specific cell surface marker is challenging and no universal cancer marker has been identified, and most likely does not exist due to the heterogeneous profile of tumours. Nevertheless the existence of cancer cell surface markers has led to the development of new possible therapies for cancer²⁵⁵.

5.3.1 Surface marker expression

Taking this into account we decided to have a look at the tumorigenic potential of cells throughout the differentiation protocol. In a first instance we looked at a range of surface markers associated with either pluripotency or tumorigenicity: TRA-1-60, TRA-1-81, SSEA5 and ROR1.

- TRA-1-60 and TR-1-81 are widely used in the field as stem cell-specific markers that are then lost during differentiation. These are thought to be a keratin sulfate proteoglycan and to have roles during development. They have also been tested as carcinoma markers, showing their potential for use as tumour markers²⁵⁶.
- ROR1 is expressed during early embryogenesis, where it contributes to organogenesis²⁵⁷⁻²⁵⁹. It is subsequently lost, and in the adult it is only detected in adipose tissue at low levels, and almost undetectable in pancreas, lung and a subset of intermediate B cells²⁶⁰⁻²⁶². It is however detected in numerous blood and solid cancers²⁶³, with increasing evidence suggesting a role of ROR1 in cancer biology.
- SSEA5 antibody was raised for detection of teratoma-forming cells and allow their removal from a population of differentiated cells²⁶⁴. Even though SSEA5 ab alone proved insufficient to completely remove the teratoma potential it proved better than classic hPSC markers and when in conjugation with other hPSC markers complete removal was achieved²⁶⁴.

A



B

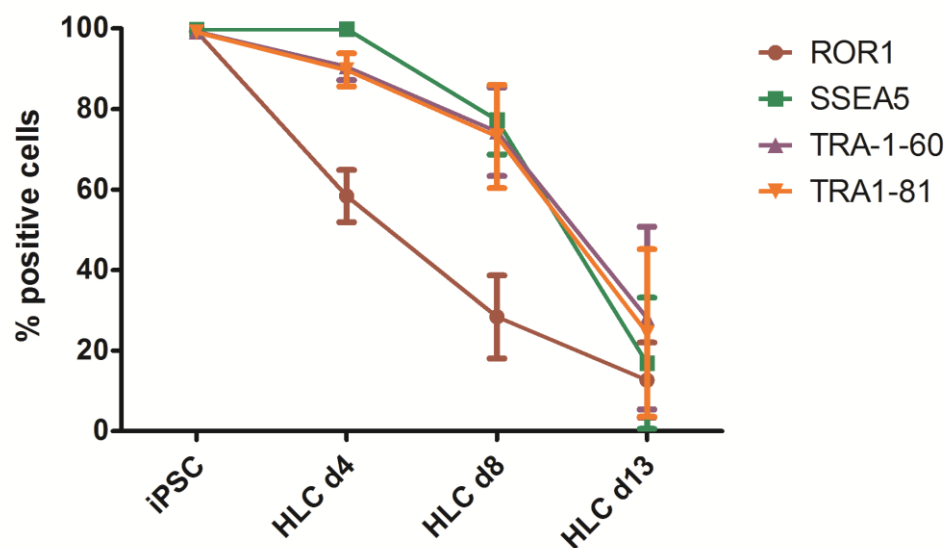


Figure 40 - Expression of surface markers assessed by FACS. Expression of different surface markers associated with tumorigenicity and pluripotency during the hepatic differentiation. A. Example of FACS profile of RDP2 iPSC line during hepatic differentiation. B. Average expression values for 3 different cell lines: RPD2, BJ and HPS1

We show that all surface markers go down during the hepatic differentiation protocol, with variability among markers and different cell lines. Despite the decrease in marker % there are still positive cells detected at day 13 of the protocol (Figure 40 and Supplementary figure 8).

A closer look shows us that ROR1 marker is the one with the most marked decrease, especially in initial stages. ROR1 has been extensively studied in CLL/blood malignancies where its use as a biomarker for CLL opened the door for therapeutics using ROR1 as a target. Besides being a biomarker, it might also provide a potential prognostic indicator as expression of ROR1 increases through the progression of CLL²⁶⁵. The same observation was seen in other blood malignancies^{260, 265-267}. In the context of breast cancer ROR1 expression levels also showed correlation with higher grade and more aggressive disease, with likely involvement in the metastatic ability of cells²⁶⁸, possible by its involvement in epithelium-mesenchymal transition (EMT)²⁶⁹.

Presence of a metastatic population in cancer is associated with later stages and more aggressive disease. CSCs have also been associated with metastatic ability²⁷⁰, and this population is crucial for cancer development when engraftment experiments were performed²⁷⁰. Removal of the CSCs showed decrease in tumour formation, pinpointing the crucial role of these cells as cancer initiating cells. Absence of this population becomes a priority when using any cellular therapy, especially for PSCs derived technologies. ROR1 has shown correlation with more aggressive stages of cancer, but whether or not it can be used to sort out the whole metastatic population still remains to be addressed.

The marked reduction of ROR1 expression during the hepatic differentiation, especially in the first stages of the protocol, can be seen as a positive feature as we might be removing the most aggressive cells and therefore reducing the risk of tumour formation significantly.

SSEA5 was raised as a marker against teratoma-forming cells. Unlike ROR1, it does not show such a decrease in expression until day 13 of the protocol. The final population is the one we are interested for cellular therapy, and therefore the one we require to be safe in a tumorigenic context. Still, the different pattern when comparing to ROR1 indicates that

ROR1 might be insufficient to assess the tumorigenic initiating population while it can still be a good prognostic for more aggressive cells.

The pluripotent surface markers (TRA-1-60 and TRA-1-81) also show a significant decrease from day 0 to day 13, which is expected as cells have committed themselves into the hepatic lineage. The genomic plasticity of PSCs gives them the ability to differentiate into multiple cellular types, but as cells differentiate this plasticity decreases to a point where their state is locked to a single somatic type. The high levels of pluripotent surface markers present at day 4 and even day 8 indicate a high plasticity of these cells. On one hand, DE cells are expected to have a high degree of plasticity as they are multipotent cells found early during development. On the other hand, cells at day 8 would be expected to have reduced plasticity as liver commitment has already occurred. This indicates presence of a population that, even though shows liver specific characteristics, still presents immature features.

The marked decrease observed between day 8 and day 13 might be a result of the passaging step. One of the possible advantages of this step was the removal of PSCs as these do not like to be in a single cell state. The data suggests that indeed this step is important in removing some of the less differentiated cells.

Overall we confirm the decrease of the pluripotent potential of cells along the differentiation protocol, also observed by the decrease in gene expression levels of pluripotent genes and upregulation of hepatic specific genes (Figure 29A, Figure 32, Supplementary figure 4 and Supplementary figure 6). This is expected as the transition from a pluripotent state to a somatic one decreases the genomic plasticity of cells.

The presence of a small population of cells positive for these markers indicates presence of cells with teratoma-initiation potential. The question remains whether this population will induce tumour formation in an *in vivo* scenario.

5.3.2 Teratoma potential

To assess the pluripotency of stem cells one of the most stringent test in the teratoma formation, by which injection of the cells into mice (routinely in the testis) lead to the formation of teratomas. If the cells are truly pluripotent the teratomas are made up of cells from the 3 different germ layers. The tumorigenic risk associated with iPSC and its derivatives does not necessarily require the cells to be pluripotent as any unorganized mass of cells that forms tumours could be dangerous for patients. By FACS we show a reduction of surface markers associated with pluripotency and tumorigenicity. But to better assess the tumorigenicity of the cells we decided to adapt the teratoma formation test used for iPSC and use it for the cells of the hepatic differentiation protocol.

Animals are injected with the cells and kept until tumours can be detected as an outgrowth, with times ranging between 6 and 8 weeks until it can be detected. As soon as tumours can be detected, mice are culled with the tumours removed for histological analysis. To allow direct comparison we kept all animals for the same period of time even if there was no evidence of tumour formation.

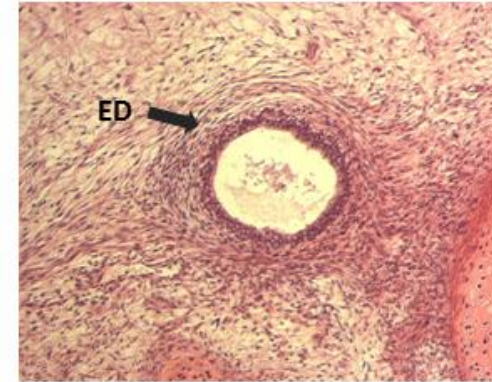
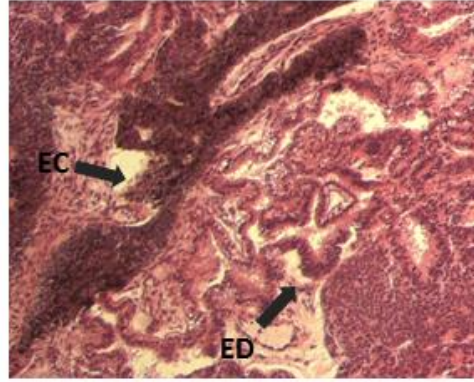
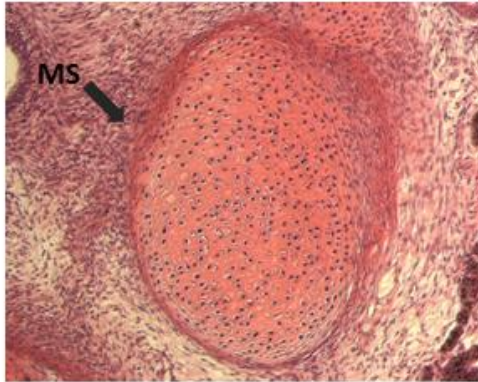
Table 5 - Adapted teratoma assay to assess tumorigenicity of cells during hepatic differentiation protocol

Number of mice	Cells injected	Outcome
2	iPSC	Teratomas with cells from 3 lineages present
2	Endodermal cells (d4 of hepatic differentiation)	Teratomas with cells from 3 lineages present. Abundance of cells from endodermal lineage
2	Hepatoblasts (d8 of hepatic differentiation)	No teratomas detected. Histologic analysis shows disruption of normal tissue architecture
2	Hepatocyte-like cells (d13 of hepatic differentiation)	No teratomas detected. Normal tissue architecture

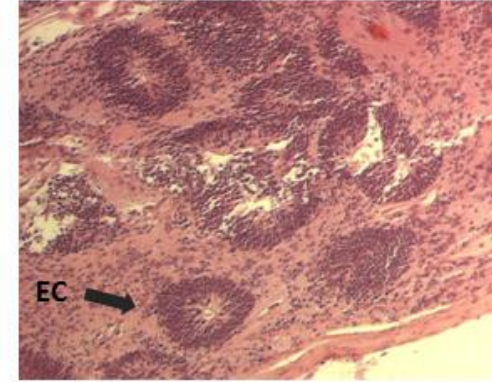
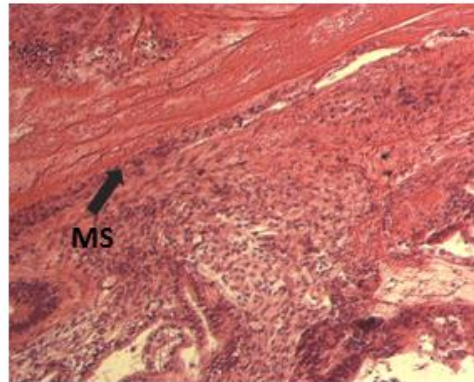
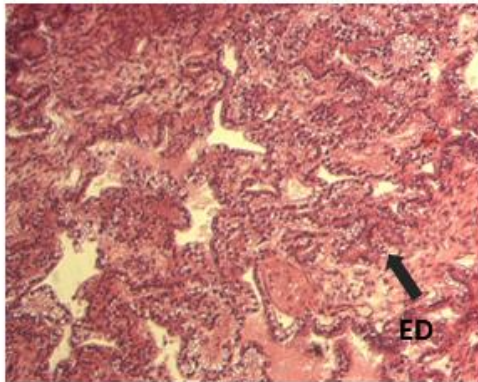
At the time point the animals were culled both iPSC and endoderm cells showed clear signs of tumour formation while testis of HLCs day 8 and day 13 showed no sign of tumour formation (Supplementary figure 9). Histological analysis of the samples provides more information on the effect of cells upon injection in the testis. As shown before, tumours of mice injected with iPSC show presence of cells from the 3 germ layers, confirming their pluripotency (Figure 41A). Mice injected with cells at day 4 of the differentiation protocol also resulted in tumour formation with cells of the three different germ layers. Cells at day 4 correspond to endodermal cells and indeed we see a prevalence of cells derived from the endoderm lineage in the tumours, but we also detect cells from mesoderm and ectoderm (Figure 41B). It is not surprising to observe cells of all germ layers as cells at day 4, even being mainly endoderm (>90% CXCR4 positive and >80% CXCR4 double positive SOX17 positive), still present very high levels of pluripotent markers (Figure 40).

Endoderm cells are multipotent and can give rise to a range of different cell types (including the liver, lung, pancreas, gut). The ability to differentiate into different cells is characteristic of stem cells, with different levels of stemness existing, dependent on the genetic plasticity of cells. The higher plasticity the less differentiated cells are, and the lower plasticity the more committed they are. This is also shown by the reprogramming process by which the higher plasticity cells possess the easier it is to reprogram them into a pluripotent state (eg: haematopoietic stem cells (HSCs) generate 300x more iPSC colonies than terminally differentiated B and T cells¹⁵).

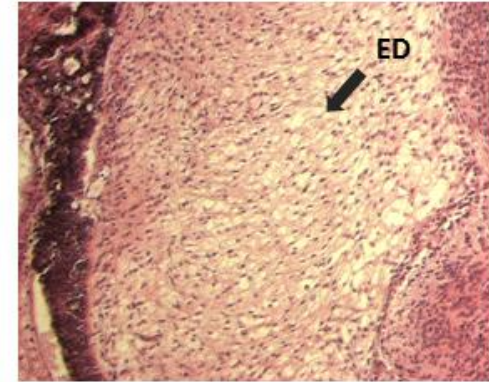
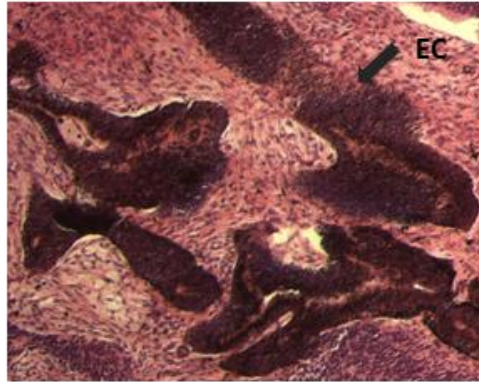
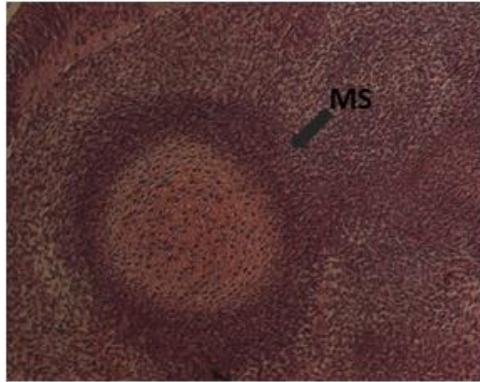
A



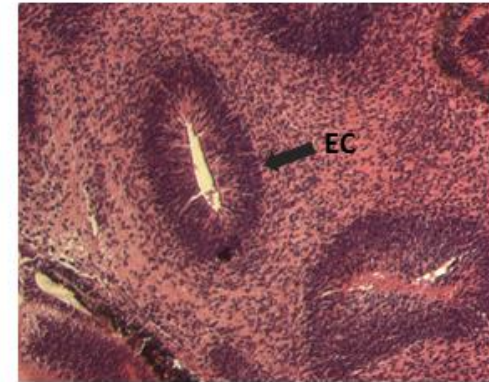
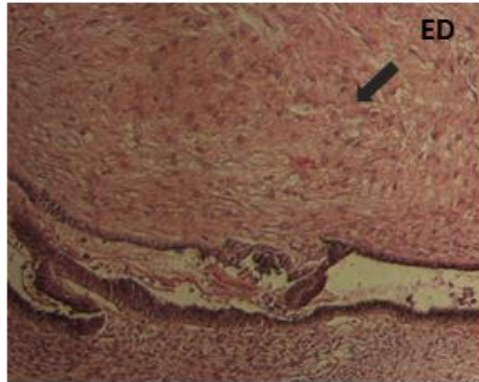
135



B



136



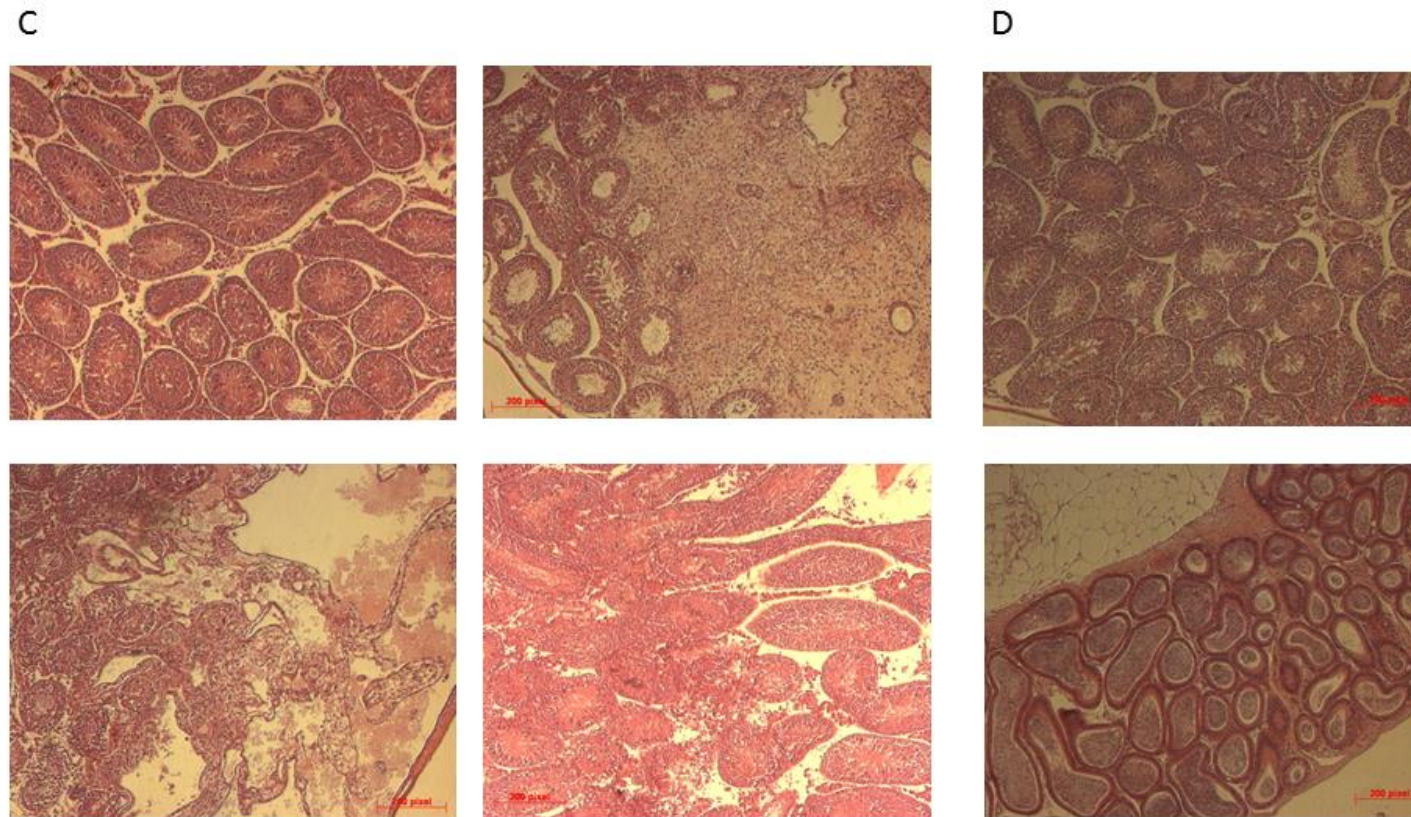


Figure 41 – Teratoma formation with cells at different stages of hepatic differentiation. Evaluation of tumorigenicity of cells during the hepatic differentiation assessed using teratoma assay. A. Histological slides from teratomas of mice injected with iPSC cells B. Histological slides from teratomas of mice injected with HLC d4 cells C. Histological slides from testis of mice injected with HLC d8 cells D. Histological slides from testis of mice injected with HLC d13 cells.

Furthermore the endodermal differentiation step is an artificial process that tries to replicate a process that is highly regulated and complex in a simple 2D process that lasts only 3 days. The short time frame and the artificial setting might be insufficient for the genetic changes to stabilize, and so when cells find themselves in the *in vivo* environment, the different signals they receive from the niche might be enough for transdifferentiation from the endodermal lineage to other lineages.

While macroscopically mice injected with HLC d8 showed no signs of tumour formation (Supplementary figure 9), histological analysis of the testis shows a slightly different story. These showed disruption of the normal tissue architecture and presence of other cell types (Figure 41C).

Of note is the fact that the testis did not show the multitude of cell types as iPSC and HLCs d4 did, rather just showing tissue disruption with cellular growth. Some endodermal structures might be present but it is not clear. Unlike cells at day 4, cells at day 8 find themselves down the hepatic lineage and so their genetic plasticity is reduced, making it harder to transdifferentiate into cells of other lineages. Finally testis of HLC d13 injected mice showed no signs of tumours or abnormal tissue architecture (Figure 41D).

Absence of tumour does not mean cells are not tumorigenic. As in cancer, where cells can have different levels of aggressiveness, cells *in vitro* can also possess different levels of tumorigenicity. This tumorigenicity model allows cells at different stages of the differentiation protocol to incubate *in vivo* for the same period of time. Same incubation periods allow a direct comparison between the tumorigenic potential of these cells but might restrict the growth potential of cells from later stages. The fact that mice injected with HLCs day 8 show disruption of normal testis cellular architecture with presence of foreign cellular structures shows that even if no tumours are yet formed, the base for it to happen is already set. FACS showed a decrease in tumorigenicity of cells as we go down the hepatic differentiation. The lower percentage of cells with the ability for tumour formation affects the rate at which tumours can arise. And so even if no signs of tumour formation were present we cannot exclude the possibility that longer incubation periods would allow for tumours to grow.

Nevertheless it is important to note that this model gives cells permissive conditions for tumour formation: very high number of cells in a matrigel matrix which provides a niche for cells to survive, aggregate and proliferate. In the context of cellular therapy there should not be such permissive conditions for tumour formation so one must be careful when interpreting these results.

Ideally cellular therapy requires a population of terminally differentiated cells that have no risk of losing their phenotype. Even in the human body – the gold standard for correct signalling and differentiation – the loss of phenotypical stability is observed. The body has regulating mechanisms to prevent survival of deregulated cells but these sometimes escape them which can result in the development of cancer.

Ex vivo generated cells might never fulfil all the safety criteria for transplantation. The saying “better be safe than sorry” is behind stem cell regulations when it comes to cellular therapy, a cautionary approach that has the patients interests at heart. Even if *in vitro* models were perfect differentiation systems, the risk of developing cancer can never be excluded, as it is a natural occurring phenomenon that results from loss of regulating mechanisms. Increased knowledge about cancer and the mechanisms for its development might help improve the stem cell field including the regulations for their therapeutic use.

A possibility that has come to light is to genetically modify cells by introducing suicide genes. This allows the removal of cells from the organisms in case any problem occurs. This option is appealing in theory but it brings up new problems such as the introduction of genetically modified cells into patients. This encompasses a whole new field that has its own ethical and legal problems and regulations. Besides, by the time a problem is recognized it might be too late for the use of a suicide gene approach. As a tumour forms a microenvironment is created that can recruit host cells, change their phenotype and incorporate them in the tumour. So even if the genetically modified cells might be responsible for tumour formation, in the end the tumour will comprise of a variety of cells, not all resulting from the transplanted cells. So the suicide gene approach might clear the genetically modified cells but clearance of the tumour is not guaranteed.

Challenges still remain regarding the use of PSCs derived cells for cellular therapy, but with the first clinical trials taking place a better idea of their safety and efficiency should soon be accomplished. But before stem cell therapies can be applied in liver disorders a better understanding of the cells is required. We know cells present some phenotypical and functional characteristics of hepatocytes but a more detailed study on one of its specific liver functions could give us better insight on their quality.

Due to the importance of coagulation for life, and the major role the liver plays in it, coagulation presents itself as a good strategy to evaluate the quality of our liver development model using iPSC.

Chapter 6 Coagulation during development

6.1 Introduction

Being one of the biggest organs in the body, the liver is responsible for a variety of functions, amongst them the production of coagulations factors important for homeostasis. Homeostasis, from the Greek *homeo* + *stasis* meaning “standing still”, is the balance between blood coagulation and fibrinolysis, securing normal blood flow in the body. In case of injury it is crucial for the prevention of blood loss and repair of damaged tissue or vasculature.

In this way homeostasis comprises the process of blood coagulation and fibrinolysis (dissolution of the blood clot). These processes are connected and need to be highly regulated to avoid either blood loss or thrombosis. Many components are part of these processes, as depicted in Table 6. These proteins depend on interactions with each other and their surroundings for proper regulation of homeostasis. Many of the proteins are serine protease enzymes that circulate in the blood as inactive zymogens waiting for an activation trigger. Upon presence of the activation trigger, often a form of vascular injury, the process of coagulation is activated with a cascade of zymogen activation takes place that culminates in repair of the damaged tissue by production of a fibrin plug.

Table 6 - Components of blood coagulation

Factors	Descriptive name	Function
Procoagulants		
I	Fibrinogen	Precursor of fibrin
II	Prothrombin	Precursor of thrombin
III	Tissue factor	Cofactor

V	Labile factor	Cofactor in the prothrombinase complex
VII	Proconvertin	Linked to TFC activates FX and FIX
VIII	Antihemophilic factor	Cofactor in tenase complex
IX	Christmas factor	Activates FX
X	Stuart-Prower factor	Converts prothrombin to thrombin
XI	Plasmathromboplastin antecedent	Activates FIX
XII	Hageman factor	Activates FXI
XIII	Fibrin-stabilizing factor	Crosslinks fibrin polymers
	Prekallikrein	activates FXII
HMWK	Kininogen	Activation cofactor for FXII and FXI, generates bradikinin
Anticoagulants		
	Antithrombin	Inactivates thrombin, FIXa, FXa, FXIa and FXIIa
	Protein C	Inactivates FVa and FVIIIa
	Protein S	Enhances protein C activity
	Heparan sulfate	Links activating antithrombin III
	Thrombomodulin	Thrombin receptor allowing linking to protein C
TPFI	Tissue factor pathway inhibitor	Inhibits TFC, FVIIa and FXa
	Heparin cofactor II	Inactivates thrombin
Fibronolytic		
	Plasminogen	Precursor of plasmin
tPA	Tissue plasminogen activator	Activates plasminogen
	Urokinase	Activates plasminogen
Antifibrinolytic		
PAI-1		Inhibits tPA
PAI-2		Inhibits tPA
	α -antiplasmin	Inactivates plasmin
TAFI	Thrombin activatable fibrinolysis inhibitor	Inhibits plasminogen activation

This view of coagulation led to the presentation of the cascade model of coagulation in the 1960s²⁷¹, also known as the classical model, as it explains the process of coagulation by which sequential activation of coagulation factors results in the formation of thrombin, that converts fibrinogen to fibrin. It proposes sequential steps by which one clotting factor leads to the activation of another, eventually leading to fibrin generation (Figure 42).

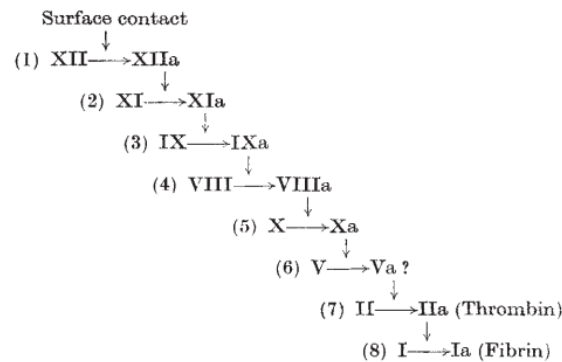


Figure 42 – Original cascade model representation²⁷¹

The original model has since been modified to include the observation that not all coagulation factors possess enzymatic activity but rather function as cofactors for proper activation of other factors. This model is often presented with two distinct activating pathways that converge at the level of Factor X (FX) in the common pathway (Figure 43). The two pathways – intrinsic and extrinsic – are often presented as distinct, independent activating pathways. Even though this presentation helps display the different interactions of the factors and how they work in a more logical and simplified way, it is too simple to explain the physiological events that occur *in vivo*. In reality the factors in the two pathways do not work independently from each other but rather interact in complex and interconnected ways with each other and the cells involved in the clotting process²⁷²⁻²⁷⁴. Despite not being a true representation of biological events, it is a good exemplification of the coagulation process in laboratory based tests.

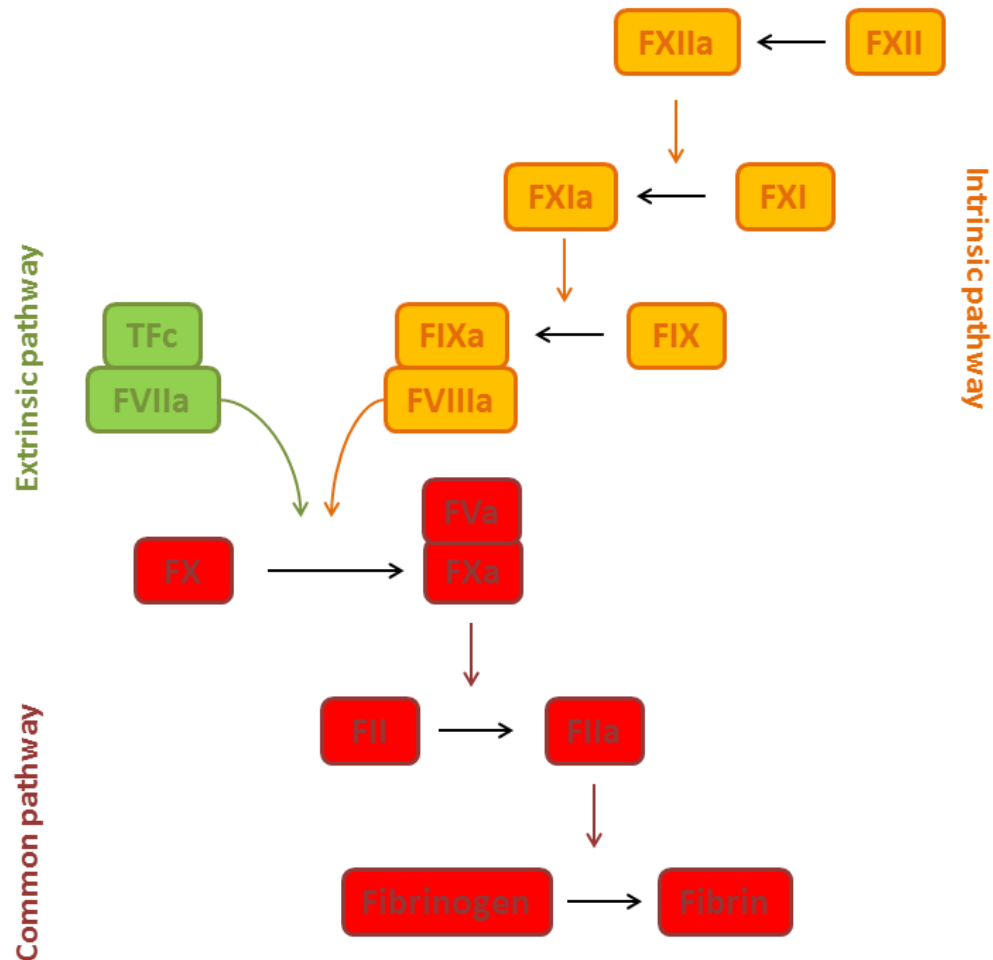


Figure 43 – Classical model of coagulation

A new model trying to incorporate the involvement of cells – cell base model – tries to present the process in a more similar way to what happens *in vivo*. Here the process of clot production is shown with interactions of cells and the different coagulations factors, derived in part from experiments that use cells as source of tissue factor (TFc) (like fibroblasts) and activated platelets as surface for thrombin generation^{275, 276}. Coagulation is presented as three different stages, which even though not independent events, can be divided as: initiation, propagation and amplification^{276, 277}.

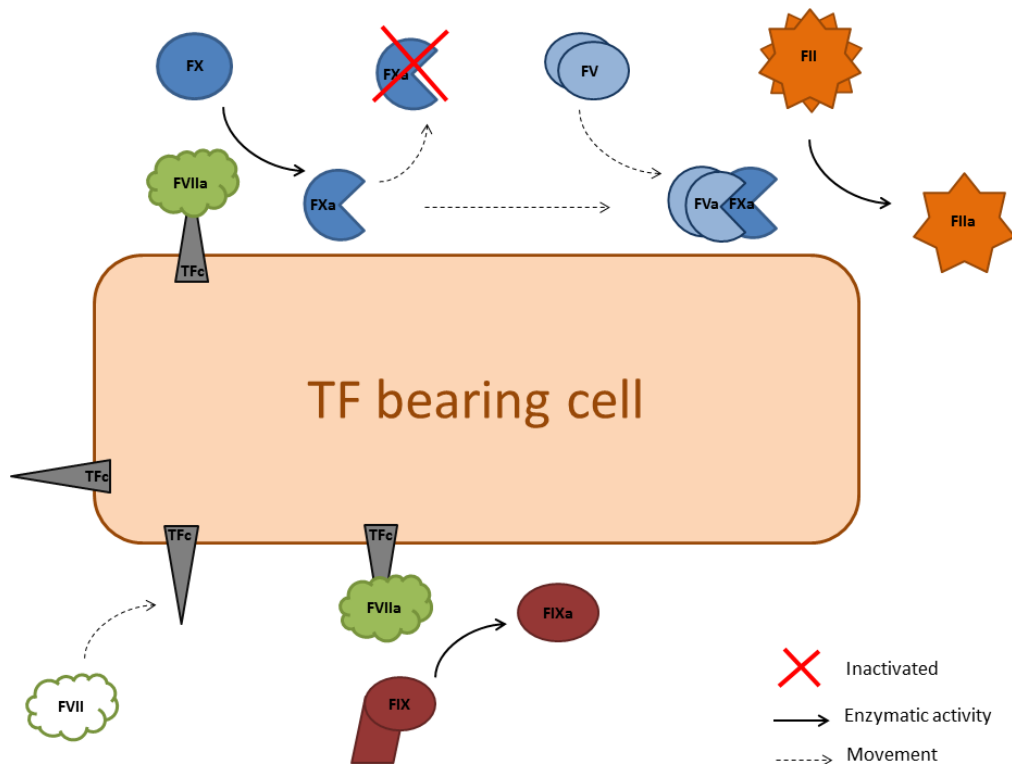


Figure 44 - Initiation step of cell based coagulation model

Coagulation is triggered when there is disruption of the vasculature. Blood crosses the disrupted site and TFc bearing cells become exposed to blood. FVII present in the blood enters in contact with TFc (its cofactor), becomes activated and there is formation of the TFc/FVIIa complex – crucial for activation of coagulation (Figure 44). This TFc/FVIIa complex at the site of injury acts on circulating FX and FIX, resulting on their active forms FXa and FIXa, respectively. FXa interacts with its cofactor FV, activating it and forming the prothrombinase complex (FXa/FVa) at the surface of TFc bearing cells. If FXa diffuses from the cell surface it is quickly inactivated by TFc pathway inhibitor or antithrombin. However the prothrombinase complex at the surface of cells acts on prothrombin, with production of small amounts of thrombin. Even though we cannot really separate the phases as it is all happening at the same time, formation of thrombin represents the end of the initiation phase. Besides the main line of events that results in the production of thrombin, during the initiation phase a loop signalling also occurs by which the activated FXa, FIXa or thrombin also act on FVII, further activating it to FVIIa.

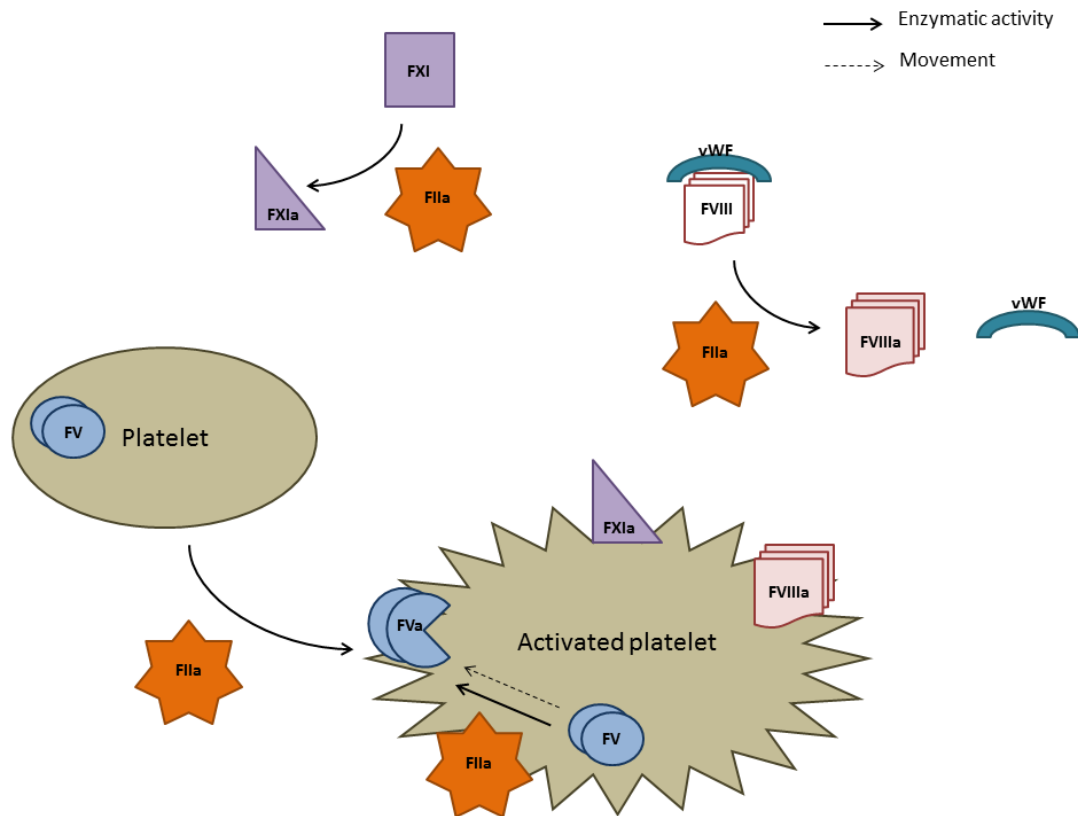


Figure 45 - Amplification step of cell based coagulation model

Production of low levels of thrombin in the initiation phase is not enough for fibrin clot formation, but is enough to initiate a burst of thrombin generation. During the amplification phase (Figure 45), the thrombin from the previous stage acts at different levels to amplify the signal for thrombin production. Activation of platelets is essential for clot formation as these are required not only for signalling purposes but also as the first level for the clotting plug. Platelets localize at the site of injury by binding to collagen or via von Willebrand factor (vWF), which partially activates them. For fully activation the thrombin produced in the previous step is required, activating the platelets via protease-activated receptors (PARs). This activation results in degranulation that releases partially active FV which is then fully activated by thrombin. Thrombin also cleaves FVIII, releasing it from vWF. Furthermore, it converts FXI at the platelets surface into its active form FXIa. By the end of this stage the activated platelets present at their surfaces all the elements for the propagation phase and the burst of thrombin production.

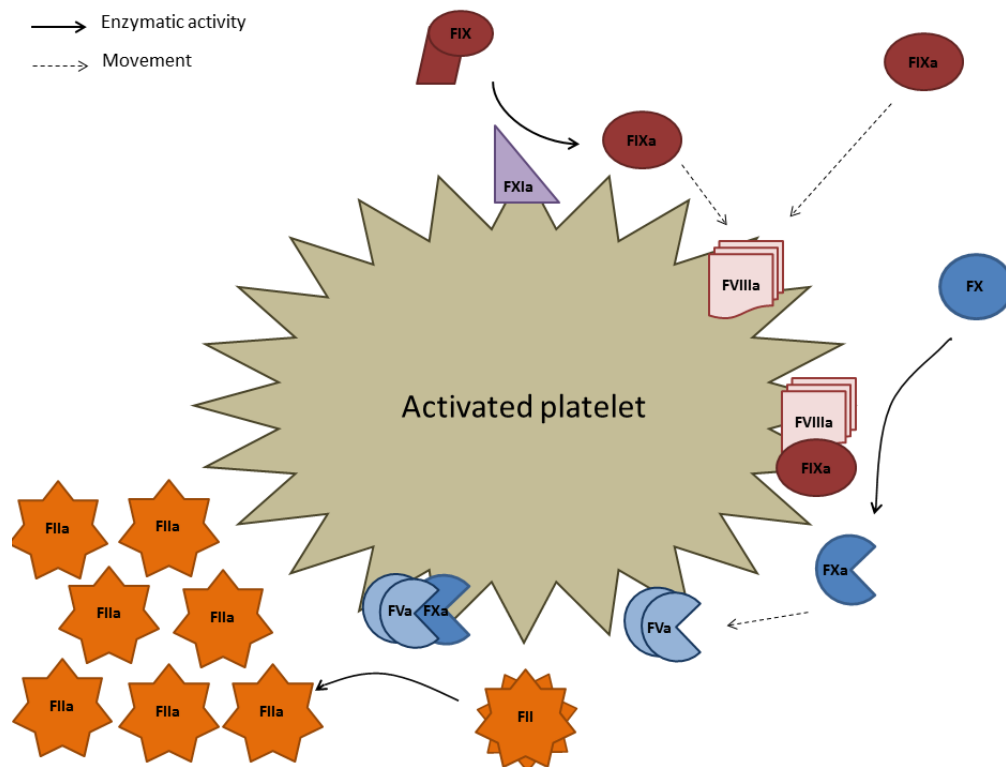


Figure 46 – Propagation step of cell based coagulation model

Propagation phase (Figure 46) follows amplification, occurring at the surface of the activated platelets recruited in the previous phase. FXIa from previous phase acts on FIX generating more FIXa, and the presence of platelets allows FIXa to roam to its surface and bind the activated FVIIIa forming the intrinsic tenase complex (FIXa /FVIIIa) on the platelet surface. This complex is crucial for the activation of FX to FXa for the production of high levels of thrombin. The production of FXa by the tenase complex at the surface of platelets is crucial since FXa produced during the initiation phase at the surface of TFC bearing cells is quickly inactivated when it diffuses away from the cells. The FXa at the surface of the platelets binds to its co-factor FVa forming the prothrombinase complex allowing for thrombin generation. It is at the surface of the platelets that enough thrombin is produced to act on fibrinogen and form fibrin that spontaneously polymerizes into strands that form the fibrin network over the platelet plug and eventually closing the wound. The thrombin produced at this stage is sufficient and required for fibrin production but it also has other functions, like the activation of FXIII to FXIIIa, activating TAFI and even being incorporated in the clot.

The damping down of the signal for fibrin production is very important as an over production of clots can lead to thrombosis, and greatly impede life. In the same way, it is also very important that the coagulation process stays localized to the site of injury. And so, there are mechanisms in place to allow the correct localization and timely termination of the coagulation process. This termination process is initiated as the fibrin clot is being formed and is crucial for correct homeostasis. Many proteins are involved in the inactivation process as can be seen in Table 6 and these can be anticoagulation, fibrinolytic or antifibrinolytic factors. Besides these proteins, it is thought platelets also play a crucial role in the damping down of the coagulation signal. There is a theory that there are different populations of activated platelets during the process of coagulation, with distinct and important functions for wound healing. The platelets refereed previously are activated and possess thrombin generating ability (Collagen And Thrombin stimulated - COAT platelets), that results from the ability of coagulation factors and co-factors to bind at its surface and leads to the eventual production of thrombin. Non-COAT platelets (without ability to produce thrombin) are thought to join at the site of injury closer to the end of the process and join the plug. The absence of coagulation signalling functions allows to damp down the pro-coagulation signal as the clot starts to close the wound.

The cell base model is a truer representation of coagulation as it happen *in vivo*, and looking back at the cascade model we can consider that the factors of the extrinsic and common pathway – TFC, FVII , FX, FV, Prothrombin and fibrinogen – could be looked at in a different perspective as the factors required for the initiation phase. In the same way, the intrinsic factors are the ones required for the signal propagation for the production of the fibrin clot. This model shows that indeed the two pathways are not redundant, with the extrinsic pathway operating on the surface of TFC-bearing cells to initiate the coagulation process, while the intrinsic pathway operates on the activated platelets surface to produce the burst of thrombin for the production of the fibrin clot.

The present cell base model also fits better with the current knowledge of bleeding disorders such as Haemophilia. Haemophilia is a disorder in which patients have defective blood

clotting system, normally the result low or absent levels of clotting factors. The most common forms are haemophilia A (FVIII deficiency) and haemophilia B (FIX deficiency). If the cascade model was accurate, the absence of either FVIII or FIX should not affect clotting as the impairment of the intrinsic pathway would in theory be compensated by the extrinsic pathway, and that is not observed. Haemophilic A and B patients bleed because the intrinsic tenase complex (FIXa/FVIIIa) is not formed properly, which impedes the activation of enough FX for the burst of thrombin production. And so, patients have a slow and inefficient formation of blood clots, with the severity of the disease being dependent on the levels of the coagulation factors in the plasma. The cell-based model is also able to explain why FXI is not essential for coagulation. It is known that patients with FXI deficiency have mild or no bleeding tendency. FXI works as an enhancer or booster for FIX activation, and in the context of this model the activated platelets allow for FIX activation even without FXI²⁷⁸.

6.1.1 Coagulation during development

While coagulation has been characterized extensively in the adult, not much is known at earlier stages, especially during development. Levels of the factors are reduced with only 10-25% of adult levels at 19-23 weeks of human foetus gestation. Levels increase progressively and at time of birth only a few have reached adult levels²⁷⁹. Based on “adult” understanding of coagulation, it would be expected that regulation of the factors involved in coagulation would be temporally coordinated, while in fact some discordant expression is observed²⁸⁰.

Indeed there is evidence indicating a preferential expression of the extrinsic pathway factors earlier during development²⁸¹.

Gene	EMBRYO DEVELOPMENTAL AGE (DPC)^								
	7.5	8.5	9.5	10.5	11.5	12.5	13.5	14.5	Liver
<i>TF</i>	+++	++	++	+++	++	++	+	+++	+
<i>FGA</i>	+++	+++	+	++	++	++	++	+++	+++
<i>FII</i>	+++	++	+	+	++	++	++	+++	+++
<i>FV</i>	++	++	++	++	++	++	++	++	+++
<i>FVII</i>	++	++	+	++	+	++	+	++	+++
<i>FVIII</i>	+	-	+++	+	++	+	+	+	++
<i>vWF</i>	+++	++	+++	+++	+++	+++	+++	+++	+++
<i>FIX</i>	-	-	-	+/-	+/-	+/-	+	+	++
<i>FX</i>	++	++	+	++	++	++	++	+	+
<i>FXII</i>	-	-	-	+/-	+	-	+/-	++	+++
<i>FXIII</i>	+/-	+	-	+/-	+	++	++	+++	+++
<i>ATIII</i>	+++	++	+	+	++	+	++	+++	+++
<i>PC</i>	+++	+++	++	++	++	+++	+++	+++	+++
<i>EPCR</i>	+++	+++	++	++	++	++	++	+++	+/-
<i>PS</i>	++	++	+++	+++	+++	+++	++	+++	+++
<i>TM</i>	+++	+++	+++	+++	+++	+++	+++	+++	+++
<i>PAR1</i>	+++	+++	+++	+++	+++	+++	+++	+++	++
<i>PAR3</i>	++	+	+	++	++	++	+	++	+
<i>TFPI</i>	+++	+++	+++	+++	+++	+++	+++	+++	+++
<i>PLG</i>	++	++	+	+	++	++	++	++	++
<i>tPA</i>	+	+	+	+	+	+	+	+	+/-
<i>uPA</i>	+++	++	++	+++	+++	+++	++	+++	+
<i>uPAR</i>	+	+	-	-	-	-	-	+	-
<i>PAI1</i>	++	+++	++	++	++	++	+/-	+++	++
<i>PAI2</i>	-	-	+	++	++	++	+/-	++	+/-
<i>α2AP</i>	+++	+++	+++	+++	+++	+++	+++	+++	+++
<i>TAFI</i>	++	+	+/-	+/-	+	+/-	+	++	+++
<i>HPRT</i>	+++	+++	+++	+++	+++	+++	+++	+++	+++

Figure 47 - Table of expression of haemostasis specific genes in mouse embryos. Copied from Ong *et al* (2000). Legend: - no signal; +, ++, +++ increasing signal strength

Ong *et al.*, in an attempt to understand haemostasis during embryonic development, characterized expression of several components of blood coagulation at different stages of mouse development. Figure 47 is a table from their paper with semi-quantitative gene expression analysis done by RT-PCR of different haemostasis-specific genes. Focusing on coagulation factors, it seems that indeed there is earlier expression of the extrinsic pathway factors. From their analysis by day 7.5 of gestation (E7.5) expression of extrinsic factors is on, while only at E10-E11.5 can most of the intrinsic factors be detected.

Moreover, Hans *et al.* showed that the levels of human coagulation factors before birth and in newborns are not identical to adults. During late stages of foetal development and the first months of life, the extrinsic pathway factors show equivalent levels of expression to the adult, while the intrinsic factors have reduced levels²⁸¹. Only from 6 months onward do all factors have levels similar to adults²⁸¹.

As described above, the extrinsic pathway factors alone cannot sustain effective coagulation in the body. For wound healing and a fibrin clot to be formed, coagulation requires more than just the extrinsic phase factors, as these by themselves do not produce enough thrombin for fibrin production (Figure 44). The thrombin produced at this stage acts on other levels such as the activation and recruitment of platelets to the site of injury for the formation of the loose platelet plug (that works as a scaffold for the fibrin network). This process sets the scene for the other stages of coagulation – activated platelets are required for the FVIIIa/FIXa complex which is fundamental in the propagation phase for production of higher amounts of thrombin and consequently fibrin, required for clot formation. And so the extrinsic factors are crucial for the initiation phase of coagulation.

And so, one hypothesis for why the extrinsic factors are turned on earlier during development, rather than all factors together, could be connected to their role in the recruitment of cells during the initiation phase, with possible involvement in vasculature formation.

“Embryonic blood vessels arise via a complex series of linked processes including the primary differentiation *in situ* of endothelial precursors (angioblasts) from mesodermal

precursors, their alignment into vascular cords to form a primary vascular plexus, and the subsequent generation of additional endothelial cells and vessels by the sprouting and splitting of pre-existing vessels²⁷⁹. This is a complex process by which extensive remodelling and maturational changes occur. Coordination of the signals and recruitment of cells for proper vessel formation is important, as well as to make sure no haemorrhage or thrombosis occurs. The presence of some coagulation factors prior to vasculature formation and the lethal phenotype associated with their inactivation²⁸²⁻²⁸⁴ (Table 7), especially the observation of defective blood vessel formation in TFc-/- mice²⁸⁴, supports the idea of embryonic involvement of coagulation factors in vasculature formation. Nevertheless the fact that not all extrinsic pathway factors result in embryonic lethality makes the role of the extrinsic pathway factors unclear²⁸⁵, with the possibility that only some factors take part in functions in vasculature formation.

Table 7 – Table with phenotypic manifestations of coagulation factors deficiency both in mouse and human

Deficiency	Mouse	Human
Fibrinogen	Bleeding associated with trauma Abnormal wound healing Abortion due to maternal bleeding	Bleeding
TFc	Embryonic lethality due to defective blood vessel formation	Unkown
FV	Embryonic lethality & Postnatal lethality due to severe spontaneous bleeding	Bleeding
FVII	Postnatal lethality due to severe spontaneous bleeding	Bleeding
FVIII	Bleeding	Bleeding
FIX	Bleeding	Bleeding

And so it is possible that these factors are there either as a requirement for small scale level of coagulation during development, or as necessary elements for the production of vasculature during development.

Moreover, there is evidence implicating some coagulation factors with other functions, such as potent mitogens²⁸⁶, role in tumorigenesis²⁸⁷, inflammatory roles^{288, 289} and even with nerve regeneration²⁹⁰. This shows us that indeed the understanding of coagulation during development is incomplete and that some factors might be involved in other aspects besides coagulation.

In terms of evolutionary biology, the hypothesis that the extrinsic pathway factors are turned on earlier during development is very interesting. Coagulation first appeared as a separate system in vertebrates, as a very simple system when compared to higher mammals. This basic system, as exemplified in lamprey (Figure 48), consists of the factors we refer to as the extrinsic pathway factors.

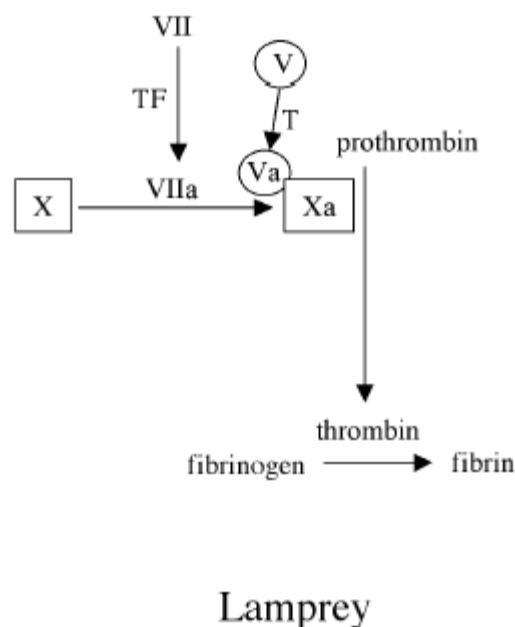


Figure 48 – Coagulation system in lamprey. Figure from Doolittle²⁹¹

Along the evolutionary road, the increased complexity of animals required for more complex systems to control haemostasis. Previous to vertebrates – the protochordates – have simple circulatory system and the clumping of circulating cells at the wound site is thought to be the only requirement in terms of a coagulation system²⁹². The lack of clotting factor genes in the genome of some of these protochordates reinforces the idea that no specialized clotting mechanism existed at this stage of evolution²⁹³.

Jawless fish are among the first vertebrates to appear, and it is with them that the invention of coagulation happens. The system is much simpler than in higher animals, but there is fibrin clot formation, produced by the action of thrombin. It is believed that thrombin precedes fibrinogen, probably with an early role in clumping cells (platelets ancestors) at the site of injury, a function it still has to this day²⁹⁴. The appearance of fibrinogen would allow a wider range of action, with a sturdier clot being formed, composed of both cells and fibrin. Duplications of the prothrombin gene would allow the appearance of coagulation factors VII and X. It is well known that many of the coagulation factors are related to each other, resulting from gene duplications, as shown by several studies²⁹⁴⁻²⁹⁶.

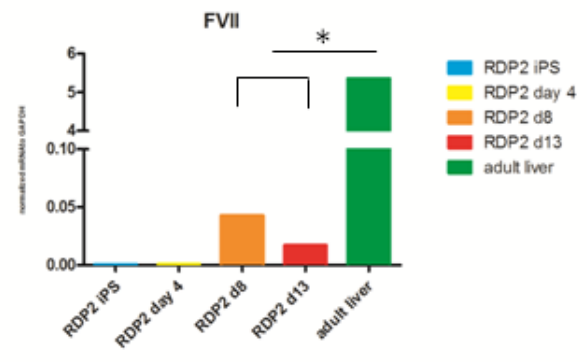
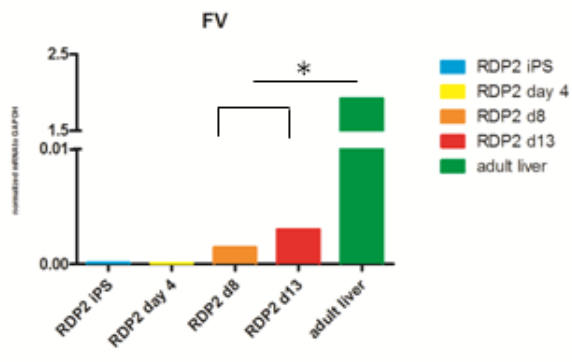
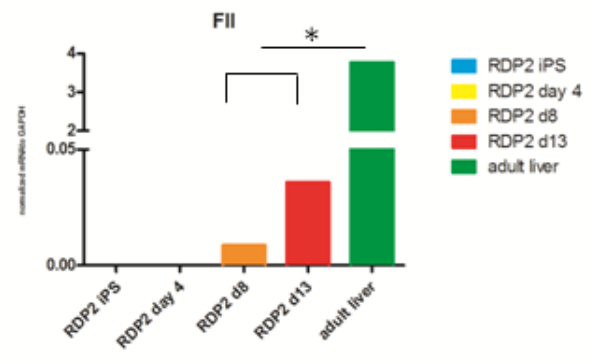
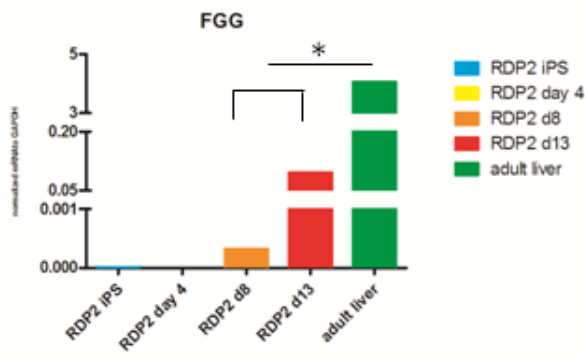
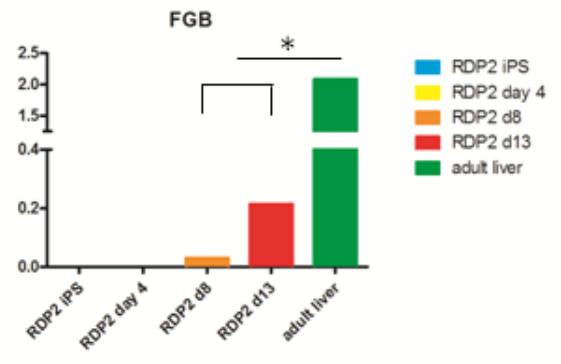
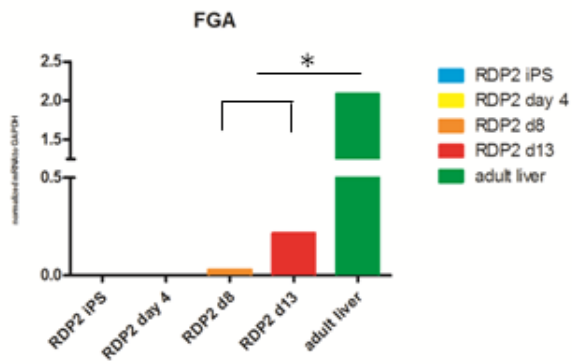
And so, the simple coagulation system observed in lampreys consisting of FX, FVII, FV, prothrombin and fibrinogen (as seen in Figure 48), introduces an evolutionary correlation to the earlier expression of the extrinsic and common coagulation factors. This could be connected to gene regulatory mechanisms that date back to the beginning of coagulation.

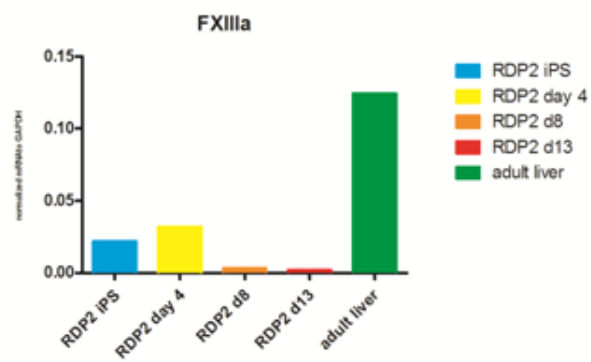
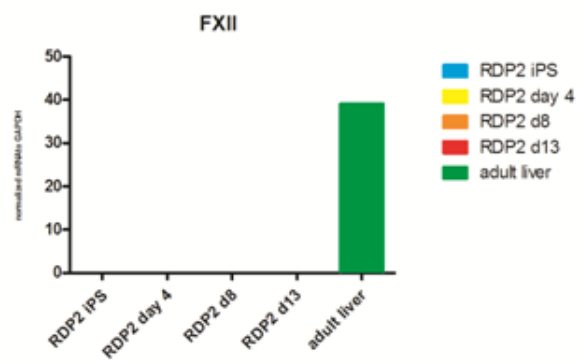
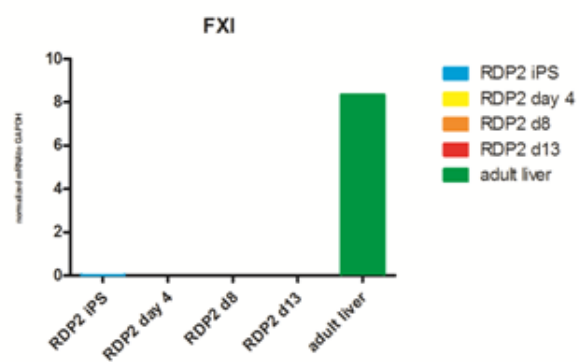
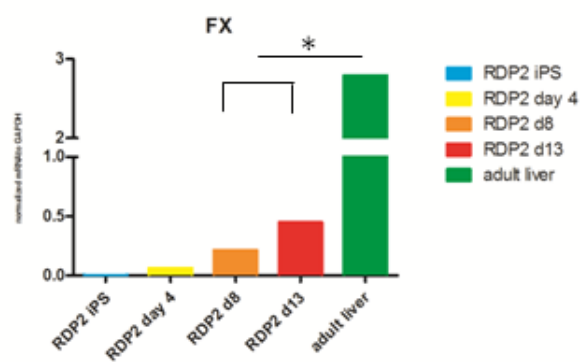
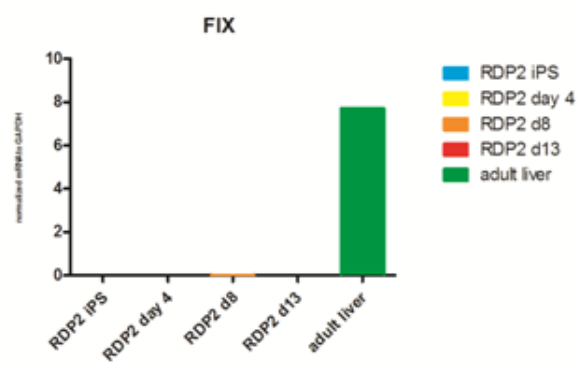
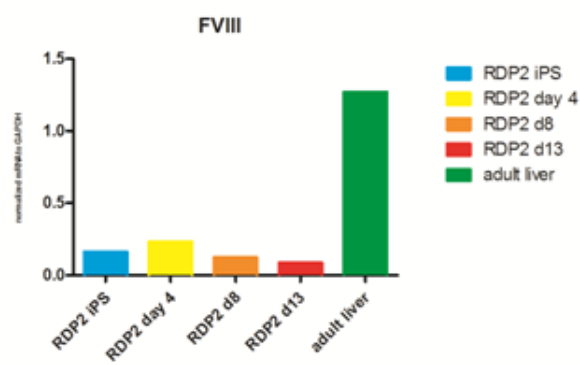
6.2 Do HLCs express coagulation factors?

Our model of liver development is focused on differentiating iPSC into HLCs. As depicted in chapter 5, HLCs from several differentiation protocols, including ours, present a foetal phenotype. While a foetal phenotype is not desirable in the context of cellular therapies, it gives us the chance to study an early developmental stage of hepatocytes in an *in vitro* setting.

Considering the hypothesis that the extrinsic pathway factors might be turned on earlier during development, this unique feature of our model allows to study coagulation during development uncover what is really happening.

And so we decided to look at expression of coagulation factors (FG, FII, FV, FVII, FVIII, FIX, FX, FXI, FXII and FXIII) at each of the stages of differentiation: day 0 (iPSC stage), day 4 (endodermal stage), day 8 (hepatic commitment stage) and day 13 (hepatic maturation stage). The relative expression is done comparing to adult liver (where all factors are highly expressed).





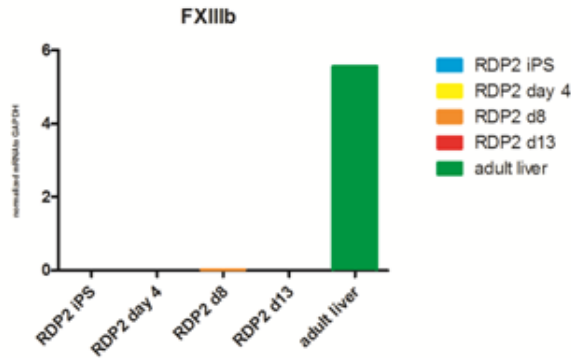


Figure 49- Coagulation factors gene expression data in RDP2 iPSC line hepatic differentiation. Gene expression by qPCR at different time points of hepatic differentiation protocol: day 0 (iPSC), day 4, day 8 and day 13. Adult liver as positive control. Expression levels normalized to GAPDH endogenous control. Stats using unpaired t test. Legend: * statistically significant ($p < 0.05$)

As shown in Figure 49, we can observe upregulation of coagulation factors from day 4 to day 8 of the hepatic differentiation protocol. This interval day4-day8 corresponds to the stage of liver commitment during development. The association between liver development and expression of coagulation factors makes sense as it is widely accepted that the liver, and primarily the hepatocytes, are the main site of production of the coagulation factors, even if some reports show production in extra hepatic sites^{297, 298}.

Not all coagulation factors are upregulated, and the factors that are detected correspond to the extrinsic and common factors: FGA, FGB, FGG, FII, FV, FVII and FX. Absence of intrinsic factors expression in the HLC model of liver development corroborates the hypothesis that the extrinsic pathway of coagulation is turned on earlier during development. As mentioned previously, this hypothesis is supported by evolutionary evidence and by the possibility of other roles these factors might have during development. If the hypothesis is true, the foetal phenotype of HLCs can justify the absence of intrinsic factors expression, that otherwise should be detected.

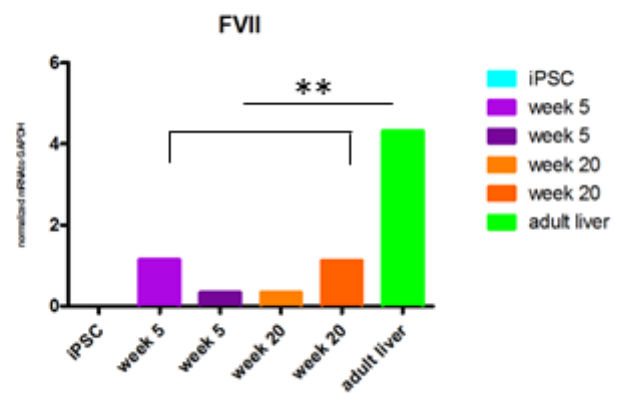
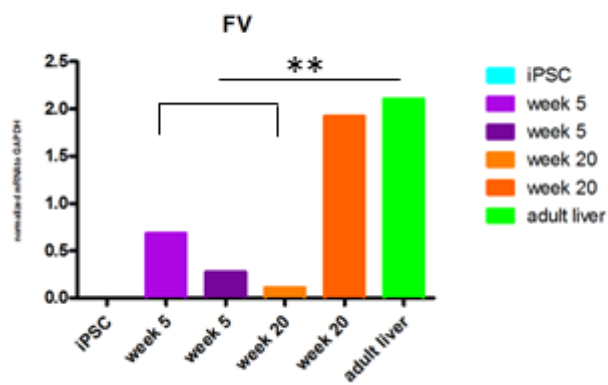
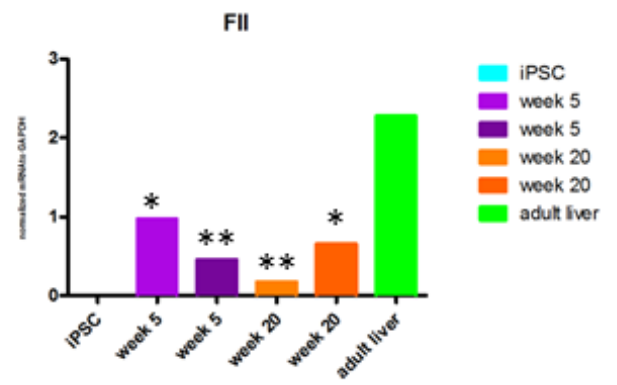
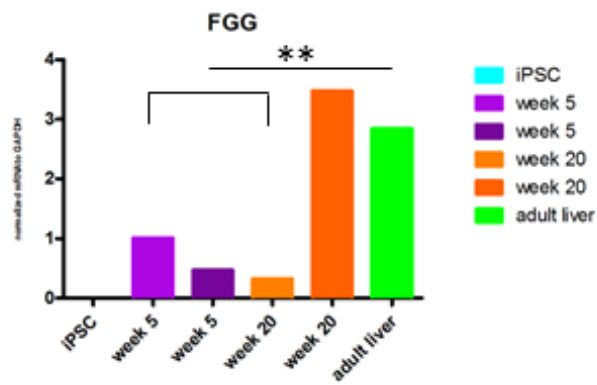
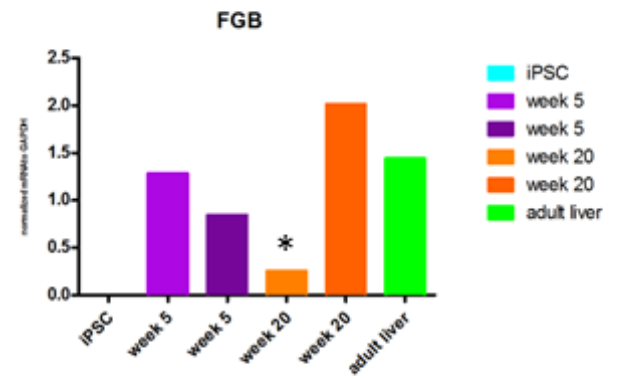
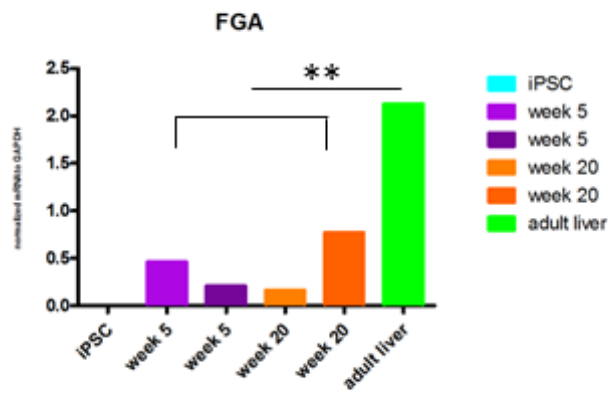
6.3 What is happening *in vivo*?

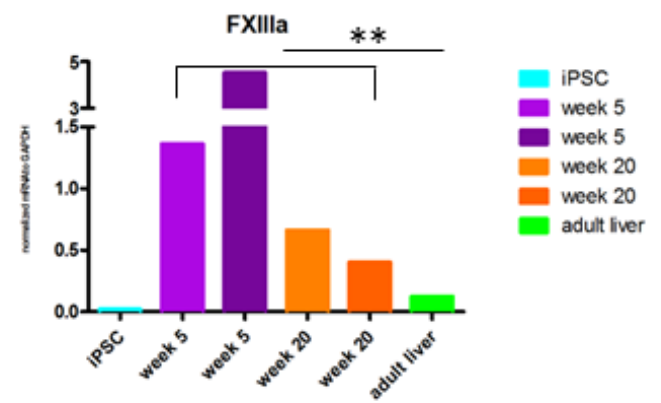
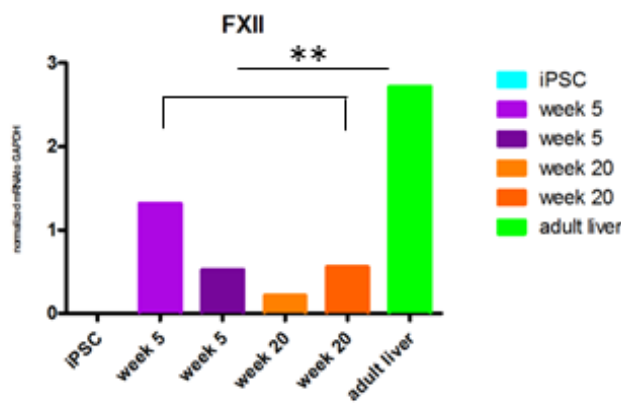
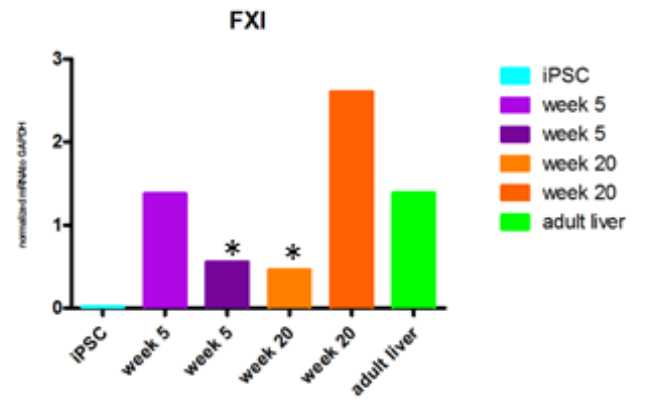
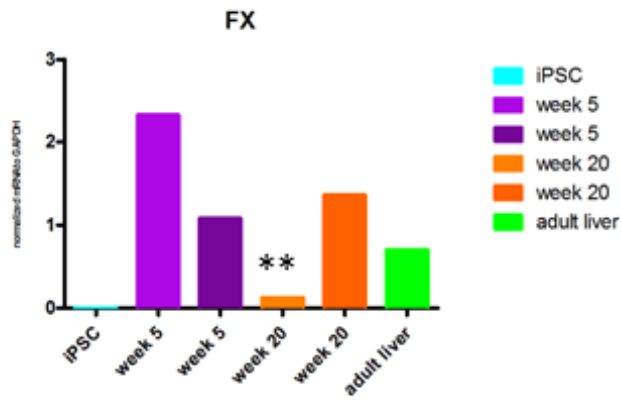
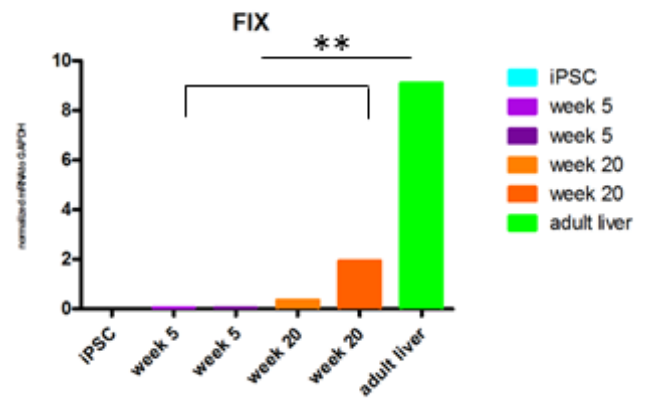
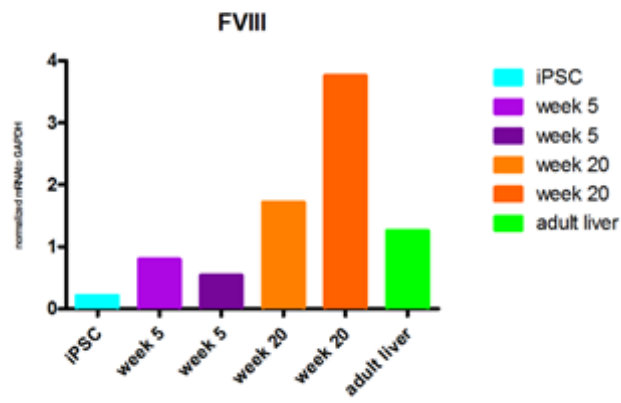
The data in HLC seems to corroborate the hypothesis that the extrinsic pathway factors are expressed earlier during development, which might indicate an important role of these factors during embryogenesis. Nevertheless it is important to ask whether our model indeed replicates the *in vivo* situation, and if so validating the hypothesis.

6.3.1 In human model

Human samples from embryonic/foetal stages are not easy to acquire as these can only be obtained from legal abortions. As a result not much has been done regarding coagulation during human development using human models. In 1990, Hassan *et al.* showed preferential expression of the extrinsic factors during human foetal development²⁸¹. They looked at gene expression by northern blot and showed that levels of expression of the extrinsic pathway factors were equivalent to adult levels while intrinsic factors, particularly FIX, showed reduced expression. Even though this study points for a preferential role of the extrinsic pathway factors during development, it also showed expression of the intrinsic factors even if at lower levels.

And so, using human foetal samples of our own (kindly provided by HDBR), we evaluated expression of coagulation factors at 2 different stages of liver development: week 5 post conception (pc) and week 20 pc.





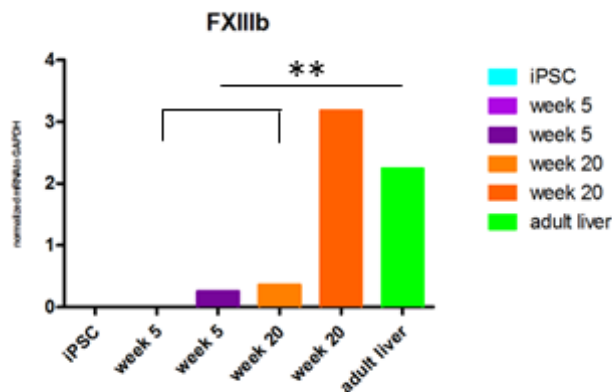


Figure 50 - Coagulation factors expression in human foetal samples. Expression levels by qPCR of coagulation genes in human foetal samples at week 5 and week 20 of gestation. Expression normalized to GAPDH endogenous control. Stats using unpaired t test. Legend: * statistically significant ($p < 0.05$) ** very statistically significant ($p < 0.01$).

As shown in Figure 50 all coagulation factors can be detected as early as week 5 pc. There is no consensus pattern that applies to all factors but in general adult liver presents significant higher levels of expression compared to earlier time points.

Both week 5 samples show similar levels of expression across the different factors, while week 20 samples show variability in some of the factors: one of the samples is closer to adult levels of expression, and the other to week5 samples. The variability observed between samples, both from same stage and different stages, is not surprising as it is known that hepatic samples show a high degree of gene expression variability. Furthermore, the samples were obtained at different times and kept frozen until requested. Different storage time and any possible differences in manipulating the samples could have affected their quality. A higher number of samples could help obtain a better trend.

In the paper by Hassan *et al.*, the extrinsic pathway of coagulation is implicated as the preferential pathway during development because the expression levels detected were equivalent to adult liver, while the intrinsic factors have lower levels of expression. Unlike Hassan *et al.* we do not see a preference to the extrinsic factors, and while *FX* and *FGB* show equivalent levels of expression to adult liver, other extrinsic factor are significantly reduced

(Figure 50). We also see some intrinsic factors with equivalent expression to adult levels (Figure 50). Their observations were based on RNA levels assessed by northern blot, a semi-quantitative method. The accuracy of northern blot can put into question their results especially since our data was obtained using qPCR, a highly sensitive method.

Of note is the fact that *FIX* has very low levels of expression in all samples. *FIX* is crucial for the tenase complex (*FVIIIa/FIXa*), and its absence is known to cause bleeding disorder – haemophilia A. Absence of *FIX* could indicate that the extrinsic pathway is the active one but presence of all the intrinsic factors makes it hard to support the claim. Another curious fact is the high levels of expression of *FXIIIa* during development.

With these results, our human foetal samples aren't able to validate the hypothesis. Earlier time points might be necessary to confirm earlier expression of extrinsic and common pathway factors, but obtaining samples earlier than week 5 is not viable.

Looking at the developmental stage of the liver of Week 5 embryos, classified at Carnegie stage 15, we see that the liver is already quite developed, with its typical architecture already present. Figure 51 shows sections of the embryo where it is possible to see that the liver makes up a significant part of the embryo, occupying most of the central plate. Presence of a mature liver structure, significant in size, makes it less surprising that as early as week 5 all coagulation factors can be detected.

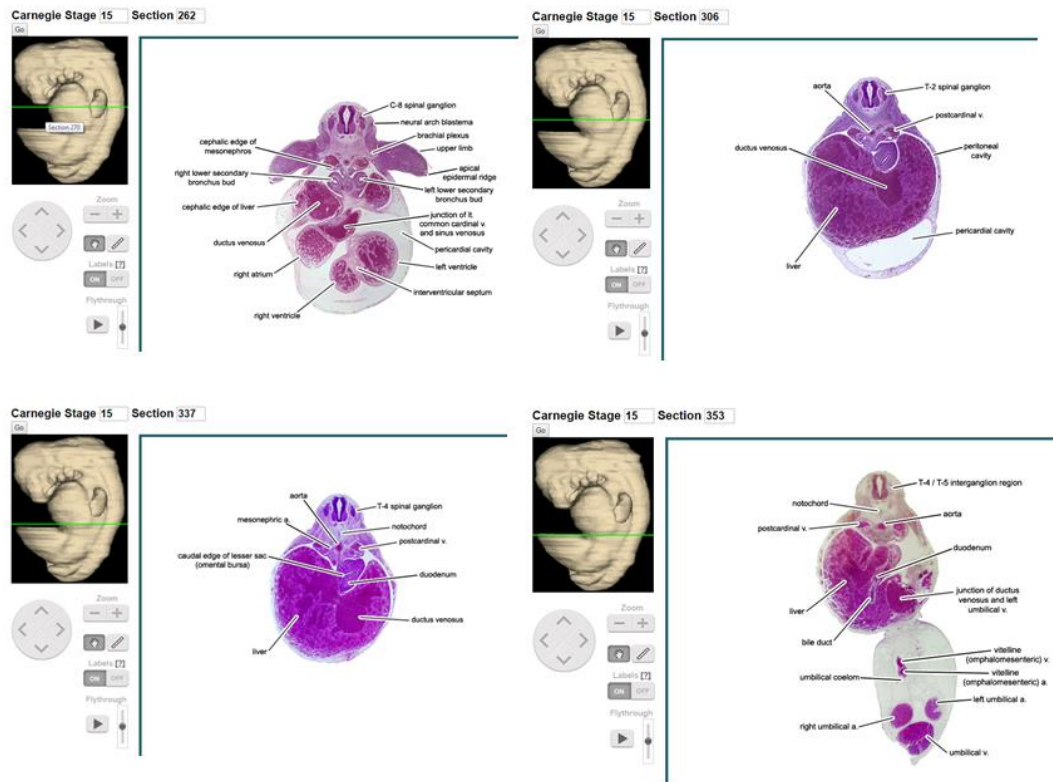


Figure 51 – Sections of human embryo at Carnegie stage 15. Liver can be observed to take a big portion of the central plate of the embryo. Figures obtained from: <http://www.ehd.org/virtual-human-embryo/stage.php?stage=15>

As earlier human samples are not possible to obtain we decided to turn to other models for the validation of our hypothesis.

6.3.2 In mouse model

Mouse is the most common model used to study human development, not only due to the fast and easy way to breed them but also for its similarities to human. Much of what is known regarding liver development was obtained from mouse studies and so it is logical to use mice as our model to assess coagulation expression patterns during development.

The study by Ong *et al.* showed earlier expression of the extrinsic coagulation factors, detected as early as E7.5 (Figure 47). Detection of coagulation factors this early in mouse development is surprising as liver development is known to start around E8.0. Liver in mice originates from the ventral foregut endoderm starting at E8.0 while the first morphological evidence is the formation of the hepatic diverticulum next to the heart at around E9.0. Bipotent hepatic specialized cells – the hepatoblasts – arise and the liver bud undergoes a period of accelerated growth between E10-E15. Maturation is gradual and occurs until after birth, with adoption of the typical architecture of the liver that allows execution of the several hepatic functions. Hence expression of coagulation factors prior to liver development is counter intuitive. Furthermore, gene inactivation studies with coagulation factors shows lethality between E9.5 and E11, which is more in line with the timeline of liver development.

To dissect what is really happening during development we decided to look at expression of coagulation factors in mouse embryos.

In a first instance we looked at coagulation factors expression in GEO datasets that used mouse embryonic samples. GEO datasets stores original submitter-supplied records (series, samples and platforms) as well as curated datasets that are publicly available, even providing tools to identify, analyse and visualize data to specific queries.

We used GEO to obtain the original gene array data containing mouse embryonic samples and analysed it using Affymetrix console with regards to coagulation factors. Datasets with whole embryo RNA were inadequate. RNA from coagulation factors is not detected, most likely due to dilution of coagulation factors mRNA in the whole embryo mRNA spectrum. GEO datasets with embryonic mouse liver RNA were available, with the earliest time point being E10.5. Going back to Figure 47 it is not clear whether the intrinsic factors are present or not at E10.5. Analysis of the data (Figure 52 and Supplementary figure 11) shows us that indeed all factors are already expressed at E10.5 and so earlier time points are required for better assessment of the expression pattern for coagulation factors.

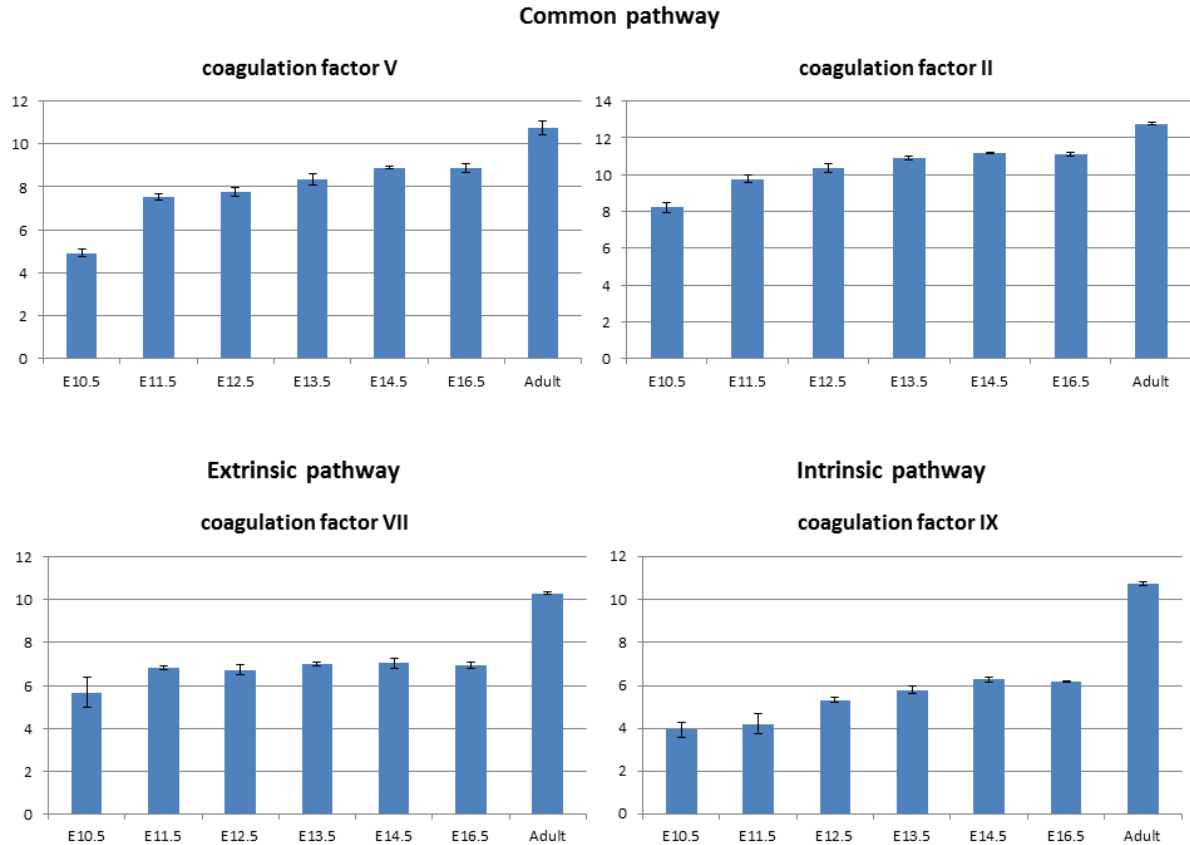


Figure 52 – Coagulation factors expression in mouse embryonic liver samples in GEO dataset. Example of coagulation factors from Intrinsic, Extrinsic and Common pathways of coagulation. Expression analysed in Transcriptome Analysis Console (TAC) from Affymetrix

To obtain mouse embryonic RNA we dissected mouse embryos and isolated liver bud structures in order to obtain a higher liver specificity and therefore increased concentration of coagulation factors mRNA. We were able to do so for embryos from E8.5 forward, as in earlier time points liver structures are not developed enough to be isolated from the embryo and so total RNA was collected (example of E7.5 and E11.5 embryos in Supplementary figure 12)

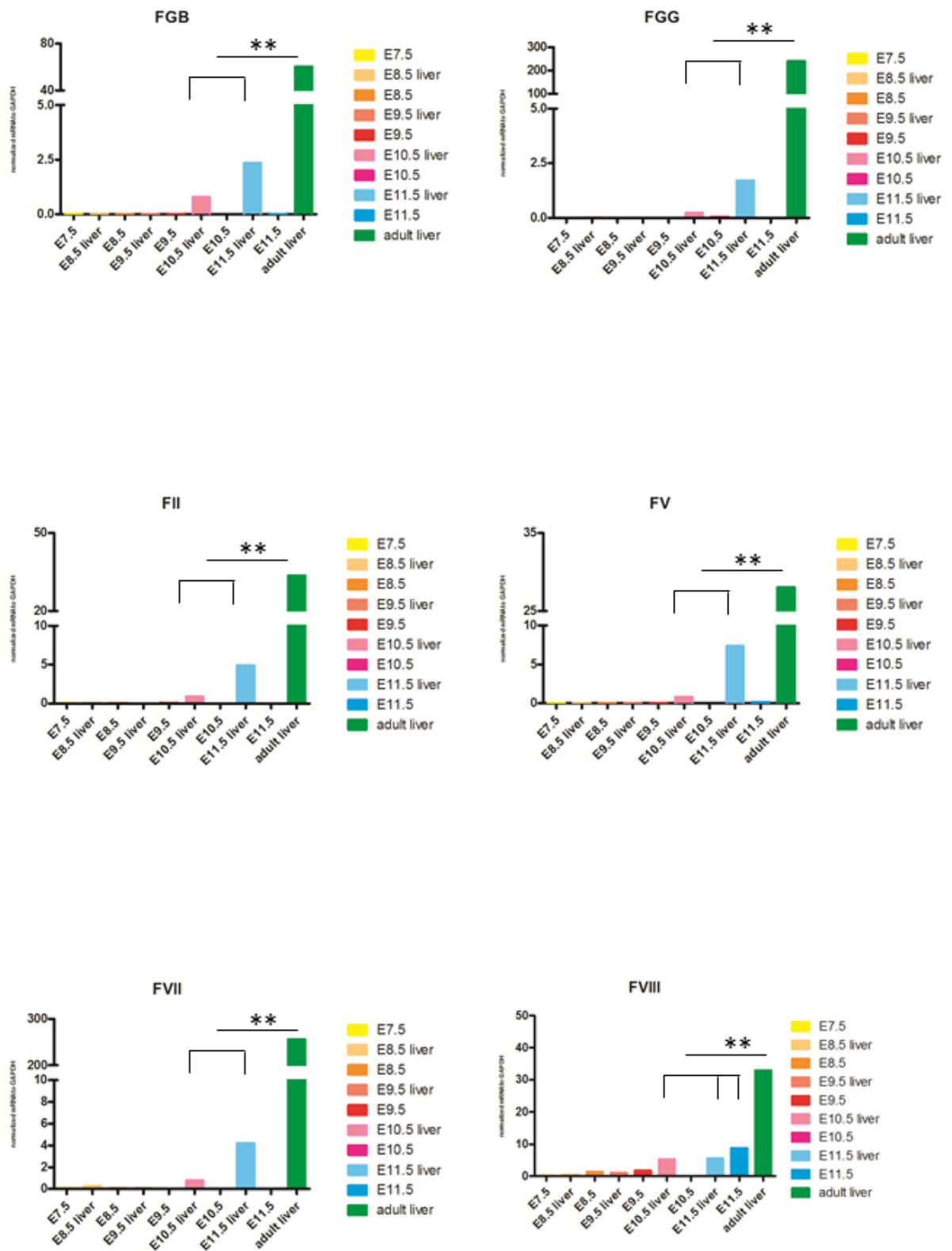
Analysis of gene expression data shows that all coagulation factors are turned on between E9.5 and E10.5 of mouse development (Figure 53). Levels of coagulation factors expression at E10.5 and E11.5 are significantly reduced when compared to adult mouse liver, which is

expected and once again disproves Hassan *et al.* theory that the extrinsic pathway factors have equivalent levels to adult²⁸¹.

Upregulation of these genes between E9.5 and E10.5 makes sense in a developmental point of view as liver structures start to emerge during this period. How were Ong *et al* able to detect expression by 7.5 is unknown²⁷⁹, especially when considering that qPCR is a more sensitive method than RT-PCR and so it should provide more reliable data.

Coagulation factor expression was only detected in the dissected liver structures of mouse embryos at E10.5 and E11.5. *FVIII* and *FXIIIa* were detected in E11.5 embryo RNA (minus liver) which can indicate extra-hepatic production of these factors during development. The exclusive expression of most coagulation factors to dissected liver structure reinforces the correct dissection of the embryos.

A curious fact is the observation that *FXIIIa* has higher expression levels during development than in the adult state. This observation is consistent across species and experiments (Figure 53, Figure 50 and Supplementary figure 11) and might implicate an important role of *FXIIIa* during development. Other factors involved in the fibrinolytic system – tissue plasminogen activator and plasminogen activator inhibitor - have shown a similar trend, with increased levels during development than in the adult²⁹⁹. The biological significance of these elevated values still remains to be clarified.



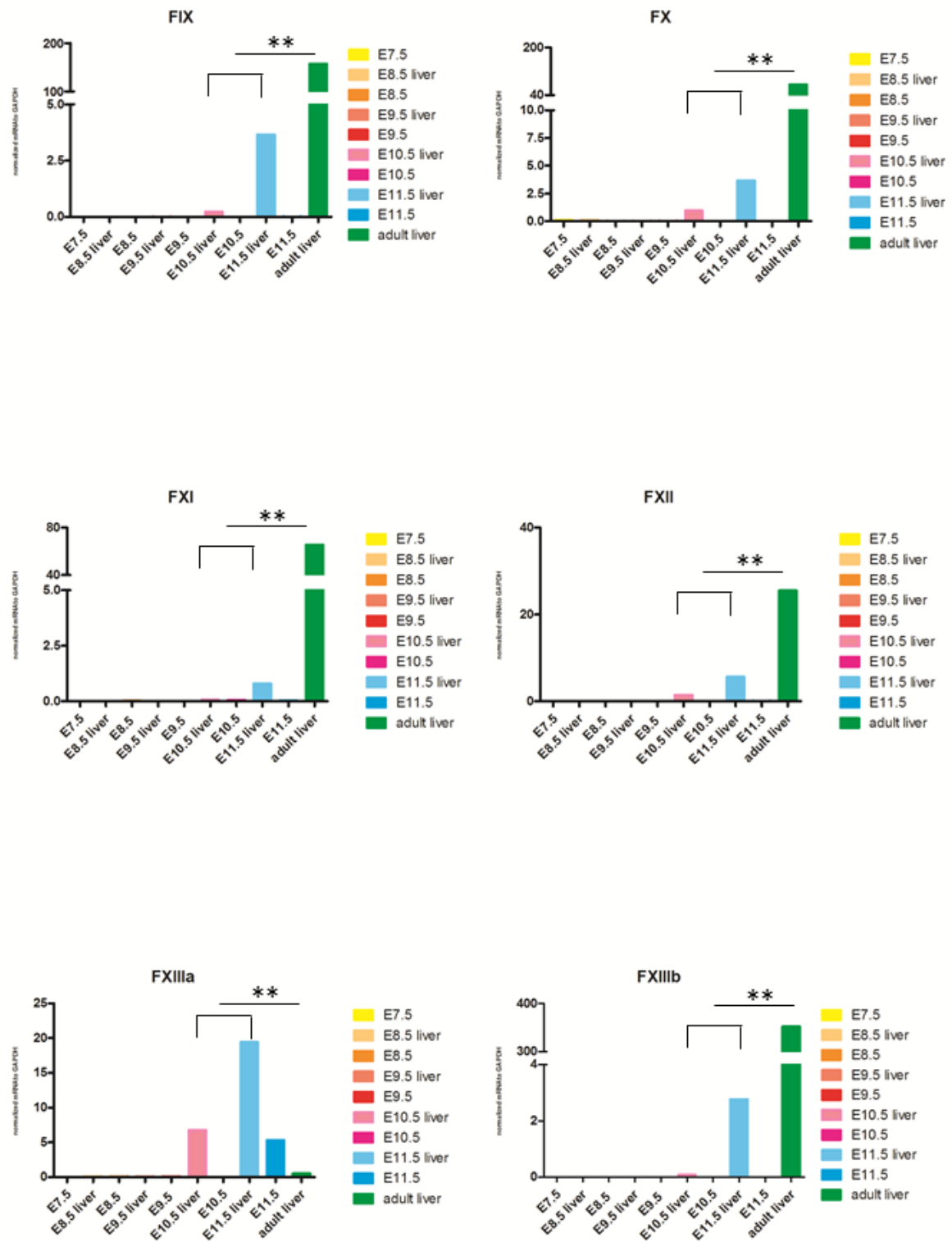


Figure 53 – Coagulation factors expression in mouse embryonic samples at different time points of development. Gene expression by qPCR of coagulation factors in mouse embryonic samples at different stages of development (E7.5, E8.5, E9.5, E10.5 and E11.5) and adult mouse liver (as positive control). Relative expression to GAPDH mRNA levels (endogenous control). Stats using unpaired t test. Legend: * statistically significant ($p < 0.05$), ** very statistically significant ($p < 0.01$).

The fact that we observe activation of all coagulation factors together indicates that there doesn't seem to be preferential expression of the extrinsic pathway. If so, the pattern of expression observed in our HLC model might just be an artefact from *in vitro* culture.

This has implication regarding the quality of HLCs. The characterization routinely done to HLCs relies on a set of tests that focus on specific liver markers and functions, but these can vary among labs. The liver is responsible for many functions, many that have never been investigated in HLCs, such as coagulation. We show here that HLCs do express coagulation factors, even if intrinsic pathway factors expression is not detected. As this appears not to recapitulate expression patterns during *in vivo* development, it can imply that HLCs have a different profile than hepatocytes when it comes to coagulation.

Quality of HLCs as a model for liver development and coagulation can be disputed, with these results able to be used on different sides of the argument: the detection of coagulation factors does show engagement of HLCs regarding coagulation but the different pattern of expression compared to *in vivo* development disputes the reliability of this engagement, a possible consequence of the effect of *in vitro* culture in the quality of hepatic differentiation. It is expected that *in vitro* differentiation can never fully replicate the *in vivo* process, with implications on the quality of the final population. To what extent does the *in vitro* setting affect the profile of cells is not known and most likely varies according to cell types and the *in vitro* systems.

6.4 HLC as a model to study coagulation

The expression pattern of coagulation factors in our HLC model is the only evidence that supports the hypothesis that the extrinsic pathway factors are turned on earlier during development. In contrast to HLC model, *in vivo* data does not support the hypothesis. Data from *in vivo* experiments has stronger credibility as it represents the natural system, while *in vitro* systems aim to replicate it but seldom are 100% successful.

This implies that the observation in HLC model could be an artefact from the *in vitro* differentiation system. Whether this observation is only seen in our model or is present in other *in vitro* differentiation system is a relevant question, as if it is model specific it implies that our differentiation protocol is not robust and should be abandoned.

One way to assess this is to use GEO datasets of other hepatic differentiation protocols and look at expression of coagulation factors. Analysis of gene array data from a different hepatic differentiation protocol²³⁶ showed no intrinsic factors of coagulation expression (Figure 54). Even though the differentiation protocol is slightly different, being longer and using slightly different cytokines, it also recapitulates *in vivo* development going through similar stages (endoderm differentiation, hepatic specification and maturation). They also characterize the cells, and not only show hepatic specific functions, but the same transcriptional profile as we assessed by RNA sequencing, therefore cells should have a similar phenotype to our HLCs. The absence of intrinsic coagulation factors indicates that this observation is not specific to our differentiation protocol and it might be a problem of the *in vitro* differentiation systems. We cannot conclusively say that all hepatic differentiation protocols show the same expression pattern but the fact that 2 distinct protocols show the same pattern strongly suggests this might be a feature seen in all *in vitro* hepatic differentiation systems.

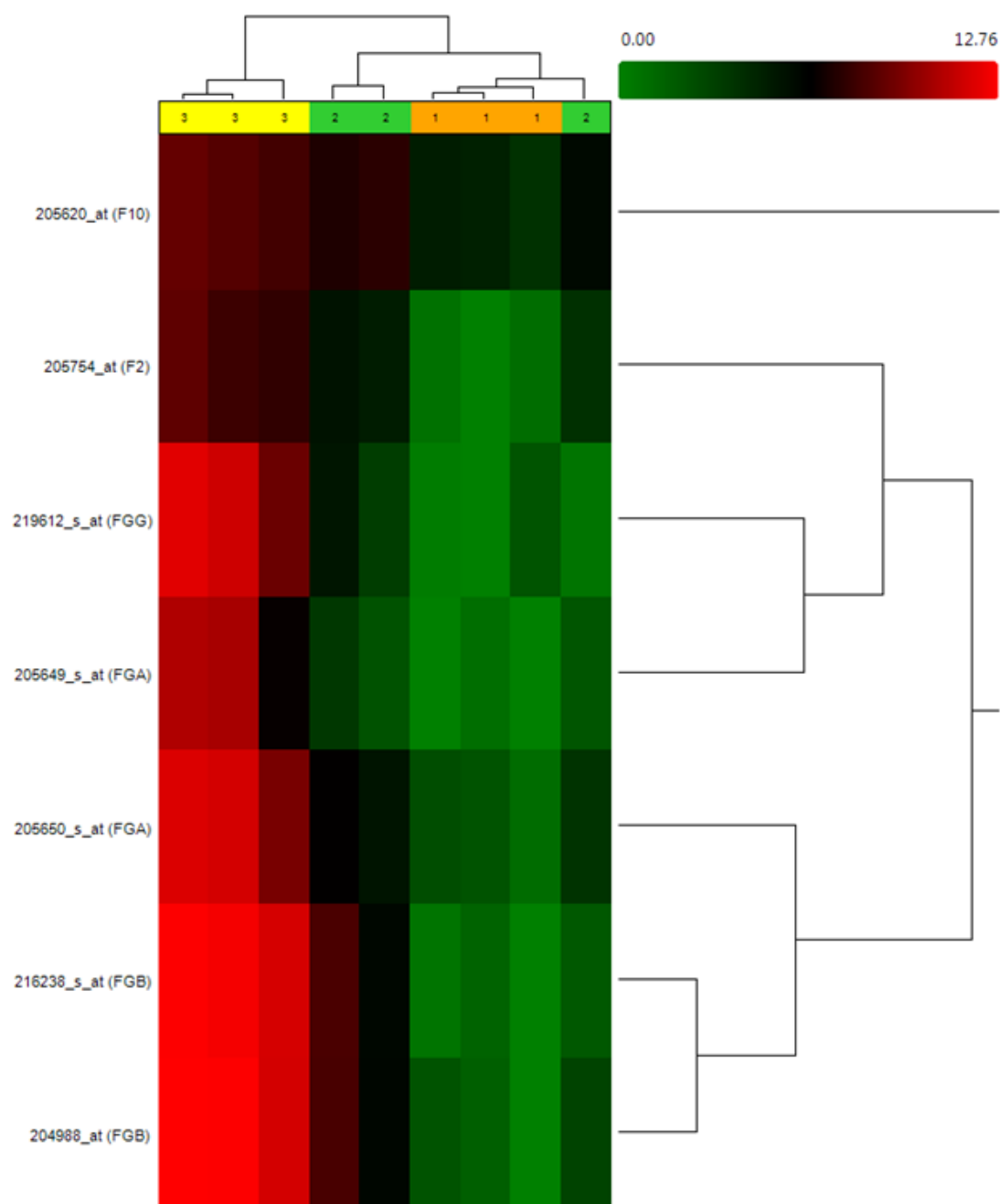


Figure 54 – Hierarchical clustering of coagulation factors expression detected in GEO dataset of cells during hepatic differentiation protocol. Output from TAC affymetrix console. Legend: 1- HLC at day5 of differentiation protocol (endodermal cells), 2- cells at day 10 (hepatoblasts), 3- cells at day 20 (HLCs). Green to red shows upregulation of gene expression between different stages. Only coagulation genes detected: FGA, FGB, FGG, FII and FX.

Even though *in vivo* data doesn't confirm the observation in HLCs, there is still a possibility that the hypothesis is indeed true. The possibility exists that the extrinsic and common pathway factors are turned on earlier in development, with the window of expression being too short to detect in the mouse model, while the *in vitro* systems prolongs or gives preference to this period and therefore we are able to detect it.

One experiment that might help shed light on the matter would be longer culture periods of HLCs to see if we can detect expression of intrinsic pathway factors. If upregulation of the intrinsic factors is observed at later time points of the differentiation protocol, a case could be made regarding the veracity of the hypothesis. Independently from supporting the hypothesis, if all coagulation factors could be detected in the *in vitro* system, HLCs would present a closer phenotype to adult hepatocytes. The problem with longer culture periods is the loss of the hepatic phenotype, as we know hepatocytes do not stand well *in vitro* conditions.

The only way to truly know what happens during human development would be to confirm using human samples. As it is impossible to obtain samples that early in development, the use of other *in vivo* models with longer developmental periods could pose an alternative.

Whether or not it is an *in vitro* artefact or real representation of what happens during development, the upregulation of coagulation factors between day 4 and day 8 of the differentiation protocols suggests that there is a common regulatory pathway being activated during this period. The next chapter will focus on this, with the aim of trying to identify the regulators of coagulation factors.

Chapter 7 Regulation of coagulation factors expression

7.1 Introduction

Transcriptional regulation is the process that controls conversion of information from DNA to RNA form. This process allows cells to continually adapt to changing conditions in their environment but is also crucial for the cell-type diversity observed in multicellular organisms.

The process of regulating gene expression is complex and multiple processes can influence expression, ranging from the chromatin state to the use of transcription factors (TFs) that bind DNA at specific regulatory sites. The genomic loci where the TFs bind are called transcription factor binding sites (TFBSs) and constitute the cis-regulatory elements. These cis-regulatory elements are classified into promoters, enhancers, silencers, and insulators among others³⁰⁰. Both promoters and enhancers are short genomic loci that serve as platforms for the binding of TFs. While promoters are usually located in the immediate vicinity upstream of genes, enhancers can be many kilobases away from the gene they regulate³⁰¹.

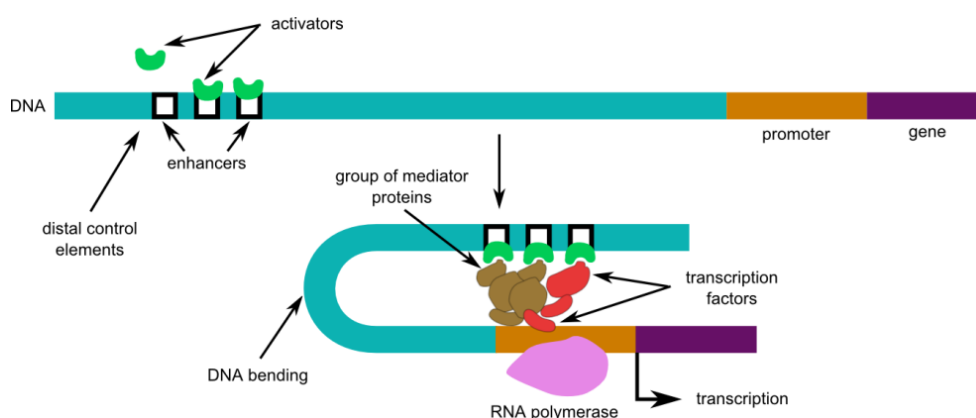


Figure 55 – Gene regulatory region where transcription is initiated

The schematic in Figure 55 is a simplified version of the eukaryotic gene regulatory region, and shows that gene expression relies on complexes formed by multiple TFs, where the rule “one transcription factor per TFBS” does not usually apply. This increases the complexity of gene expression regulation but also helps reduce the number of elements required during transcriptional regulation of genes within a cell, as each of these elements can act on multiple networks and have several functions. In this way, a TF can be involved in the expression of multiple genes, a central idea in the concept of common regulatory pathways. It is also possible that some regulatory pathways are controlled by master regulators, TFs responsible for regulating the specific network by acting on other regulatory elements. Usually the concept of master regulator is used in the context of lineage specification as the TF(s) required and sufficient for programming a specific cell fate³⁰². The same concept can be applied at a smaller scale, in smaller and clustered regulatory networks instead of the entire cellular phenotypes and lineage specification. While master regulators have been detected for certain pathways³⁰³, it does not imply every pathway will have a master regulator.

The transcriptome is the full repertoire of transcripts expressed in a cell at a given developmental stage and condition. Identifying the transcriptome is essential for understanding cellular mechanisms such as the ones that govern developmental and disease states. Human genome-wide expression analysis using next generation sequencing, such as RNA-sequencing (RNA-seq) data, provides an opportunity for in depth profiling of fundamental biological activity such as gene expression and regulation.

RNA-seq is a DNA sequencing technology that provides a high throughput method for cDNA sequencing, generating information about mRNA content and quantifying gene expression. This technology has the added advantage of reducing background noise, being un-biased (does not require a control genome) and sensitive enough to detect wider range of the transcriptome³⁰⁴, especially when compared to other available technologies such as gene arrays. RNA-seq has clear advantages over existing methods and is helping to better understand the transcriptome of various species, including in humans³⁰⁴⁻³⁰⁶.

7.2 In the search for a master regulator

The observation that all common and extrinsic coagulation factors are upregulated between day 4 and day 8 of the differentiation protocol (chapter 6 section 6.2) led us to the hypothesis that there is a common set of TFs responsible for the upregulation of these extrinsic and common pathway coagulation factors. It is likely that these TFs are upregulated prior to day 8. We, therefore, decided to search for possible master regulator(s) of these coagulation genes using RNA-seq, motif and network analysis in order to improve our understanding of the mechanism(s) regulating the so-called extrinsic and common pathway. This pathway is preserved through evolution and is critical for the initiation of blood clot formation.

We carried out a combination of *in vitro* and *in silico* work in search of candidate master regulators. Using RNA-seq, we looked for differential expression of TFs at different time points of the hepatic differentiation of iPS cells. We then performed an *in silico* exercise to identify putative transcription binding sites in the upstream region of the coagulation genes. Additionally, we searched the published databases for known protein-protein interactions. This allowed us to narrow down the TFs that may be involved in regulating the extrinsic and common pathway genes.

7.2.1 RNA-seq data analysis

RNA-seq analysis was carried out as described in methods (Chapter 2 section 2.14). In brief, RNA was harvested in triplicates from iPS cells, and differentiated cells at day 4, 8 and 13. For comparison, we also harvested RNA from primary human liver hepatocytes (Life Technologies). Our collaborators in Cambridge (Professor Willem Ouwehand, Dr. Mattia Frontini, Dr. Myrto Kostadima and Dr. Romina Petersen) performed RNA sequencing on these samples and assisted with parts of the initial analysis. The RNA-Seq read counts were imported into DESeq2 and their variance stabilizing transformation (VST) was quantified. The log fold change was plotted against the mean gene expression (Figure 56). The plot in

Figure 56 shows the unshrunk log2 fold changes from the four different time points over the mean of normalized counts (i.e. the average of counts normalized by size factors).

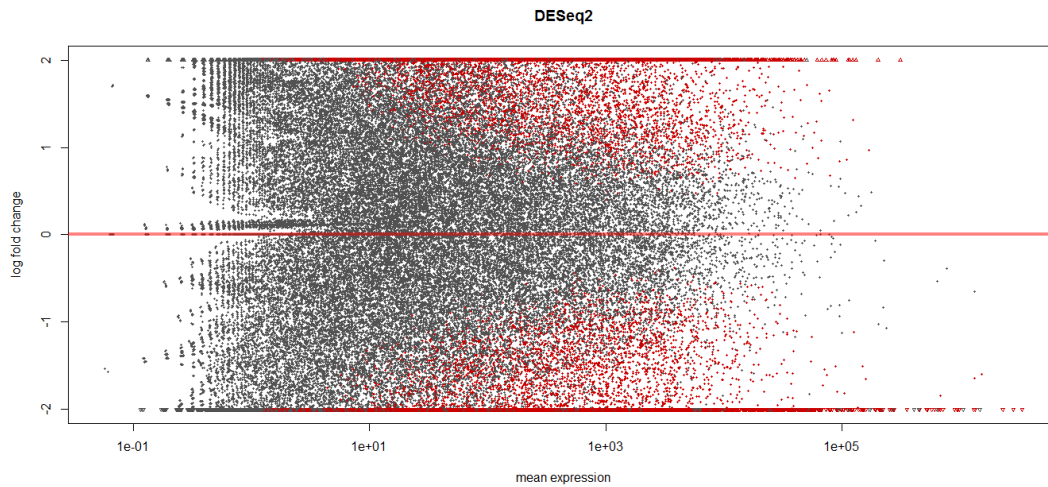


Figure 56 – Mean average (MA) plot representation of differential expression of genes in the sample overall (using DESeq2). The red dots are differentially expressed while black dots are not.

Because fold changes are non-additive and asymmetrically distributed, all of our RNA-seq analysis uses log (fold change) scale. The log (fold change) are normalized before any further analysis, to enable comparison between different time points and conditions. The log (fold change), quantity that measures a gene's expression, was calculated from the fragments per kilobase of transcript per million fragments mapped (FPKM) after normalization. The relative abundances of transcripts are described in terms of the expected biological objects (fragments or total counts) observed from a RNA-seq experiment.

Following the differential expression analysis using MMDIFF (refer to Chapter 2 section 2.14.3), only those with a posterior probability $p > 0.5$ were considered, allowing to greatly reduce the number of differentially expressed genes from the pool of potential regulators of coagulation (Table 8). Genes were further categorized based on up- and down-regulation (Table 8), but we narrowed our focus on TF that were up- or down-regulated since we were interested in factors that control expression of the coagulation genes.

Differential expression was performed between the sequential time points of the hepatic differentiation protocol – d0 (iPS cells) vs d4, d4 vs d8 and d8 vs 13 – allowing us to assess what happens between each stage. An additional analysis looked at difference between iPSCs and day 4 Vs day 8 and day 13.

Table 8 - List of differentially expressed genes between different time points in the study.

p>0.5	d0d4	d4d8	d8d13	d0d4 vs d8d13
Expressed genes	1024	1673	665	4766
Upregulated	620	849	493	2334
Downregulated	377	787	155	2302
TF	100	173	68	565
Upregulated TFs	69	98	53	293
Downregulated TFs	31	75	15	272
Coagulation	0	5	2	7

P*: posterior probability – the probability of assigning observations to groups given the data. Meaning, the probability of assigning up- or down-regulation with specific fold between two time points to a gene, given the data being analysed. The revised probability of the gene in considering being up- or down- regulated between two different time points given that the gene has equal chances of being either up- or down- regulated.

Table 8 shows upregulation of the seven coagulation factors mRNA between d0d4 and d8d13, which is in concordance with our observations in a separate technical analysis using qPCR as described in the results section of Chapter 6 (section 6.2). In terms of their $\log_2(\text{FPKM}+1)$ values, a quantitative measure of expression, Figure 57 shows the normalized log mean expression values of each of the extrinsic and common coagulation genes at each stage of the differentiation protocol. The intrinsic coagulation factor genes were either not detected or at low levels, representative of background rather than expression. This may suggest that further differentiation in these iPS cultures may be required for upregulation of all coagulation factors.

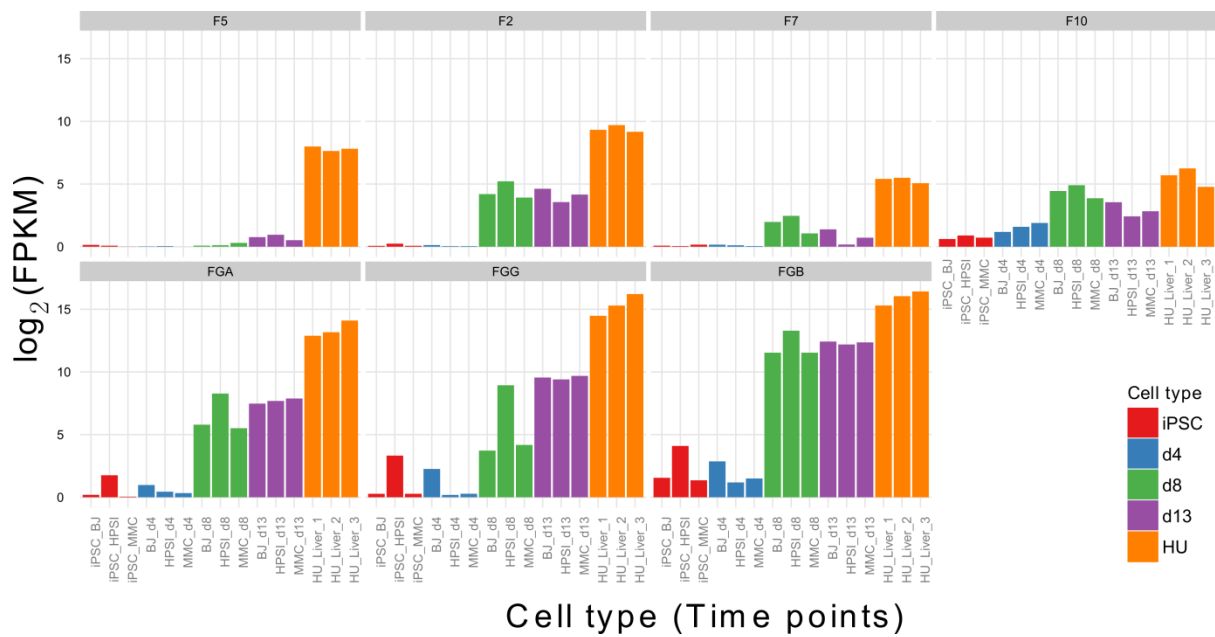


Figure 57 –Quantitative measure of expression of the extrinsic and common coagulation genes at the different stages of the hepatic differentiation protocol and in liver control. Expression levels assessed by RNA sequencing

Figure 58 shows the fold change in expression of the coagulation genes during the hepatic differentiation. Again, it confirms the upregulation of the common (*FX*, *FII*, *FGA*, *FGB*, *FGG*) and extrinsic coagulation factors (*FV* and *FVII*) occurring between day0/day4 (d0d4) and day8/day13 (d8d13). A negative value implies upregulation between two groups, meaning the gene is upregulated in the second group compared to the first. For example, *FV*, between d0d4 and d8d13 has a log₂(FPKM+1) value of -1.6 which means that the gene is upregulated in the second group (d8d13) with respect to the first (d0d4). Individual analysis between each stage was also done, confirming upregulation of the extrinsic and common factors between day 4 and day 8.

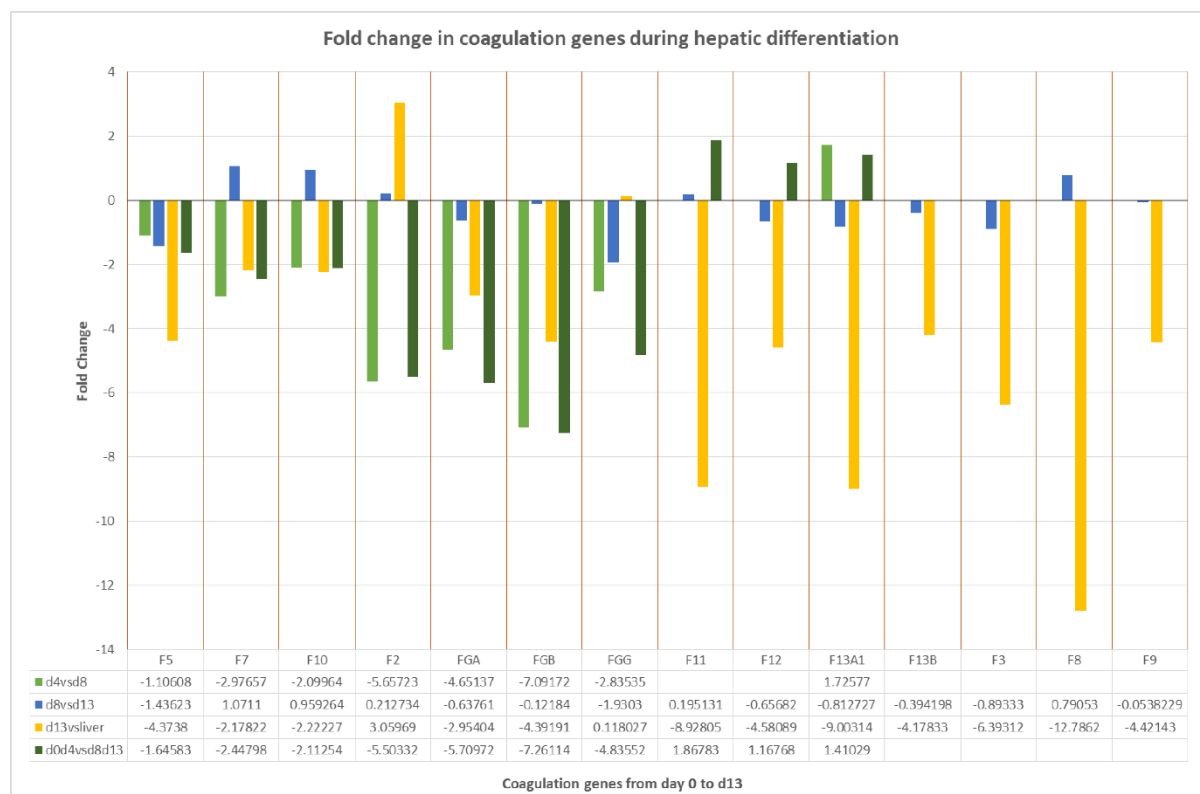


Figure 58 – Bar chart showing the fold change in all the 14 coagulation genes. Fold change shown for different groups: Day 4 Vs Day8, Day 8 Vs Day 13, Day 13 Vs Liver

As we are interested in the regulatory network for the coagulation genes, the period when coagulation genes are upregulated becomes the relevant time window to look for transcription factors. From the total 565 significantly up- and down-regulated transcription factors throughout the whole differentiation (Table 8), it is likely that the master regulator candidates are upregulated either along with the coagulation genes or before their appearance.

To identify the relevant TFs in the coagulation gene expression network we decided to use *in silico* analysis of the upstream regions of the coagulation genes to identify consensus sites for TFs (motif analysis) and crosslink this with RNA-seq data. At this juncture, the master regulatory candidates could be any from the 167 transcription factors up-regulated between day 0 and day 8 (Table 8) in the RNA-seq analysis. We also keep a track of the TFs between day 8 and day 13 to check if we could find any possible TF of the other coagulation factors that would be expressed later during development.

7.2.2 Motif Analysis

Motif analysis can help reduce the number of candidate TFs for the coagulation gene expression. Motifs are defined as a sequence pattern that is widespread and thought to have a biological significance. Using bioinformatics, motif analysis allows us to use knowledge of the DNA sequence of genes and upstream regions to look for consensus binding sites where possible regulatory elements (such as transcription factors) can bind.

Considering that the upstream region of genes normally consists of non-coding DNA, the knowledge that functional elements (such as TFBSs) are known to evolve at a slower rate than non-functional elements makes conserved DNA sites good candidates for TFBSs. This method for TFBS discovery uses alignment scripts to find conserved regions, and normally takes into consideration sequence conservation of these TFBSs

Coagulation genes are conserved between each other and are considered to have multiplied from a small number of core genes. Despite their evolutionary conservation, the promoter sequences do not share high homology. They only share 19% sequence similarity between the 15kb upstream region of extrinsic and common coagulation factor genes. Despite the low homology, they do share common TFBS motifs.

Using in-house databases (Supplementary table 1), we queried the 15kb upstream regions of our genes of interest to find TFBS motifs and identify the possible TFs that bind to them (for more detailed description refer to Chapter 9 section 9.3). From the list of all possible TFs that bind at various upstream motif regions of every coagulation gene, we filtered out duplicates and selected only those TFs that appear to bind to more than one coagulation factor. This because we are searching for master regulator(s), and by definition these will regulate multiple coagulation factors. Therefore, from the hundreds of transcription factors predicted to be binding to the 14 coagulation genes, only 95 possible transcription factors were common to more than one coagulation factor gene.

Not all of the 95 TFs regulate the coagulation genes. Hence, we carried out over-representation analysis (ORA) on these 95 potential master regulator candidates. ORA allows to check which of the 95 TFs are significantly enriched and associated with the coagulation

pathway. ORA reduced the number to 17 possible candidates, and with further filtering for liver specific TFs (in house database in Supplementary table 1), the number was reduced to 9 possible candidates, found in Table 9.

Table 9 - Master regulator candidates from Motif analysis

C/EBP
APOA4
GATA-3
GR
HNF1B
HNF4A1
NF-kappaB
OCT1

Of the 273 possible TFs differentially regulated between day 0 and day 8 of the hepatic differentiation, we filtered five master regulator candidates identified by the *in silico* motif analysis. Combining the results of both RNA-seq analysis and sequence based motif analysis allowed us to shortlist five potential candidates for the regulation of coagulation: HNF4A, CEBPZ, HNF1B, GATA-3 and NF-KappaB1 (Table 10).

Table 10 – Master regulator candidates from combining RNA-seq and Motif analysis

Master TF candidate	Posterior Probability	log2(FPKM+1)		Fold Change*	description
		day0/day4	day8/day13		
HNF4A1	0.98	1.17	9.08	-3.14	hepatocyte nuclear factor 4, alpha
HNF1B	0.99	1.63	25.58	-2.84	HNF1 homeobox B
GATA3	0.99	1.45	12.44	-2.59	GATA binding protein 3
C/EBPz	0.93	149.51	53.12	0.99	CCAAT/enhancer binding protein (C/EBP), zeta
NFKB1	0.54	1.66	3.12	-0.89	nuclear factor of kappa light polypeptide gene enhancer in B-cells

7.2.3 Gene network/Pathway analysis

Gene network/pathway analysis is another strategy that allows us to independently shortlist a set of possible master regulators for coagulation factors gene expression. To build this network for coagulation genes, the system biology approach was used. For such an analysis information from protein-protein interactions (PPI) networks (STRING and IntAct databases) as well as from pathway databases (KEGG and REACTOME databases) were used. From the human PPI, we generated a sub network with only those genes that are seen during RNA-seq liver differentiation analysis. Using this network as a template, we constructed multiple levels of regulatory networks starting from the 14 coagulation genes, to narrow down a set of possible master regulator candidates. Figure 59 represents the constructed network from the theoretical knowledge obtained from the different databases.

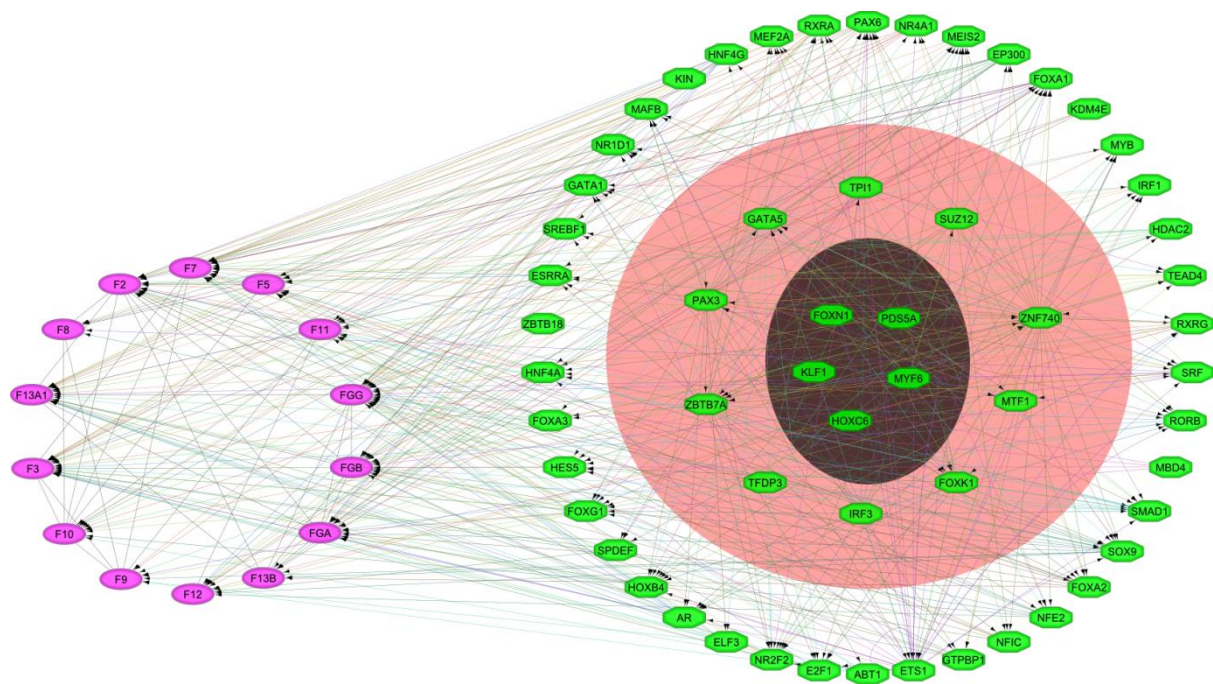


Figure 59 – Gene network for coagulation genes. Knowledge based

Combining information from RNA-seq with the previous network, we were able to remove all the TFs that are not expressed during our hepatic differentiation and a new multilevel regulatory network was constructed allowing to further narrow down the number of possible master regulator for the coagulation genes (Figure 60).

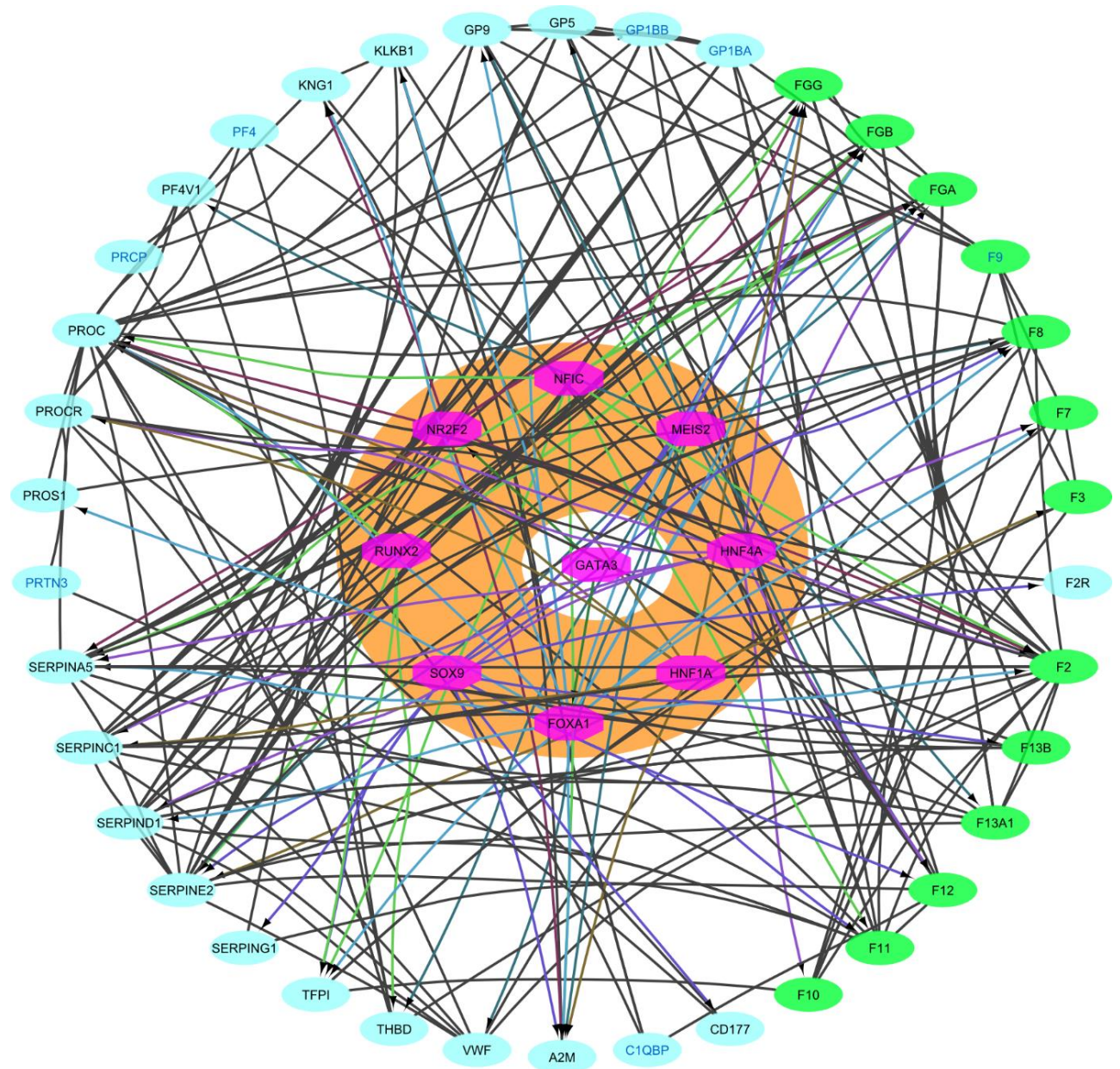


Figure 60 – Gene network for coagulation genes. Knowledge based information combined with RNA-seq data

Figure 60 shows a representation of the complex network obtained using this approach. In green, lies the starting point for this network: the coagulation factors. Several other proteins interact with them but only the factors in pink fit as possible regulators of the network.

RNA-seq together with *in silico* approaches allowed us to shortlist candidates that are possible master regulators of the coagulation network. From the possible master regulators, we find three common TFs from two independent analyses: HNF4a, HNF1 and GATA3. The recurrence of these TFs from independent analysis makes them strong contestants for master regulator of the coagulation expression pathway.

Both HNF4a and HNF1 are important TFs for liver phenotype, and while GATA3 is also known to be expressed in hepatocytes, it is most known for regulating T-cell development and in endothelial cell biology. HNF4a has previously been implicated for its role in regulating coagulations genes^{307, 308} as well as regulating HNF1 expression in hepatocytes³⁰⁹. Together with its important role in the hepatic phenotype, HNF4a was regarded as the leading candidate as master regulator of coagulation factors expression. With this in mind, we decided to examine its role in coagulation factors expression.

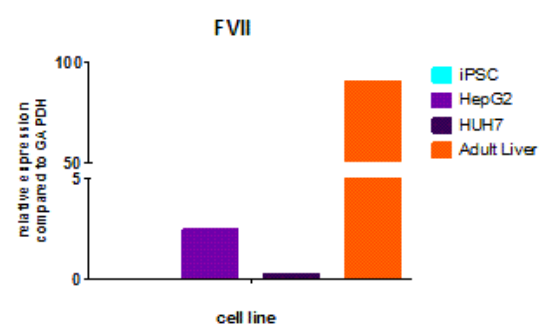
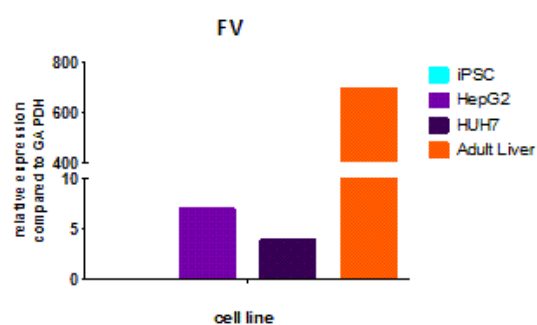
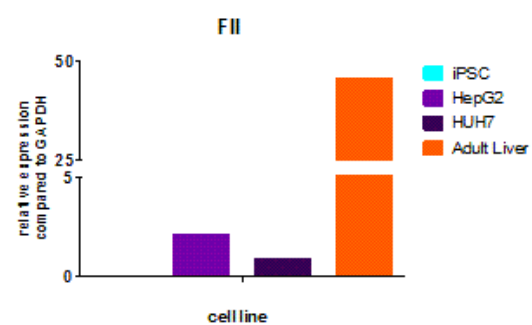
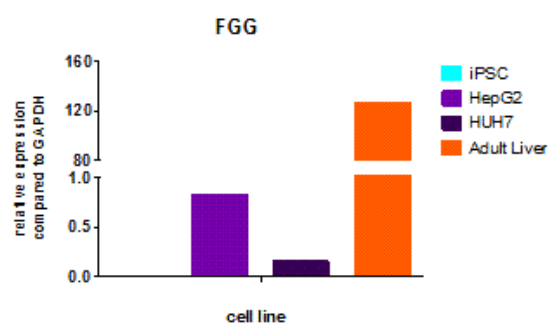
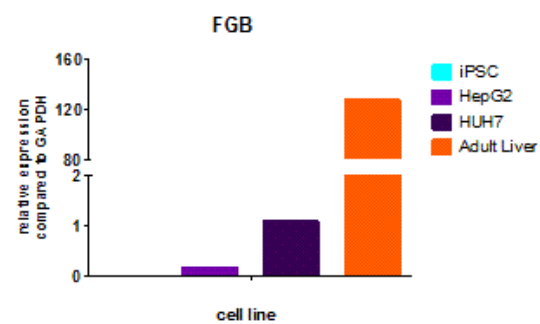
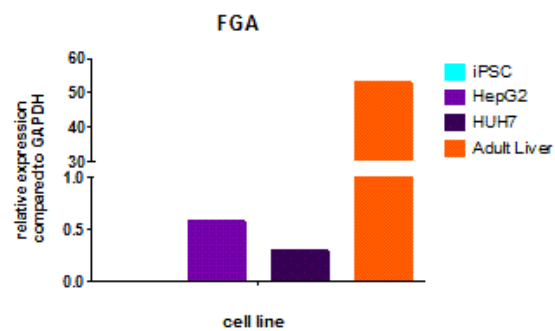
7.3 Does HNF4a regulate expression of coagulation factors?

HNF4a is a key regulator of liver homeostasis, shown to be involved in functions such as in lipid and bile acid synthesis, gluconeogenesis, amino acid metabolism, and even in blood coagulation. Knock out models have shown the importance of HNF4a for proper liver development during embryogenesis^{310, 311} and in adult disruptions of HNF4a expression has been associated with development of hepatocellular carcinoma^{312, 313}. From a biological point of view, the strong relation of HNF4a with normal liver phenotype supports the view of HNF4a as the best candidate for master regulator of the coagulation pathway.

Indeed, many of the coagulations factors are known to have HNF4a as an important TF for their expression (rulai.cshl.edu/LSPD/), but whether or not it is crucial for activation of all factors' expression still needs to be determined.

To assess the role of HNF4a in moderating coagulation factor gene expression, we decided to silence HNF4a expression in two liver cells lines: HepG2 and HUH7. These are cancer cell lines derived from human hepatocellular carcinomas (from 15 year old and 57 year old patients, respectively), and usually used as *in vitro* models for the study of human hepatocytes. Being cancer derived lines, they lack a healthy hepatic phenotype, but the use of hepatocytes is restricted by their availability and difficulty to maintain in culture. The use of cell lines has its limitations, but overcomes the problems of using our HLC model. The HLC model is a differentiation where cells go from a pluripotent state to hepatocyte-like cells. During the 13 day period, the cells undergo a series of changes, many not clearly understood, which would make it difficult to quantify the exact effect of HNF4a knock down on the cells. The fact that they are a heterogeneous population that is still acquiring the hepatic phenotype would introduce additional variables on a knockdown experiment. Therefore, the HepG2 and HuH7 cell lines provide a simpler system for the initial evaluation of the effect of RNA HNF4a knockdown.

By using two distinct cell lines, we are also ensuring that the effect of HNF4a knockdown is reproducible and not just line specific. HUH7 and HepG2 are two cell lines routinely used as *in vitro* models to study the liver. Even though they are known to possess liver functions, an important first step is to make sure they express coagulation genes. Indeed, when we checked by qPCR both lines expressed coagulation genes (Figure 61), which allows us to use them to study HNF4a as the possible master regulator for coagulation gene expression network.



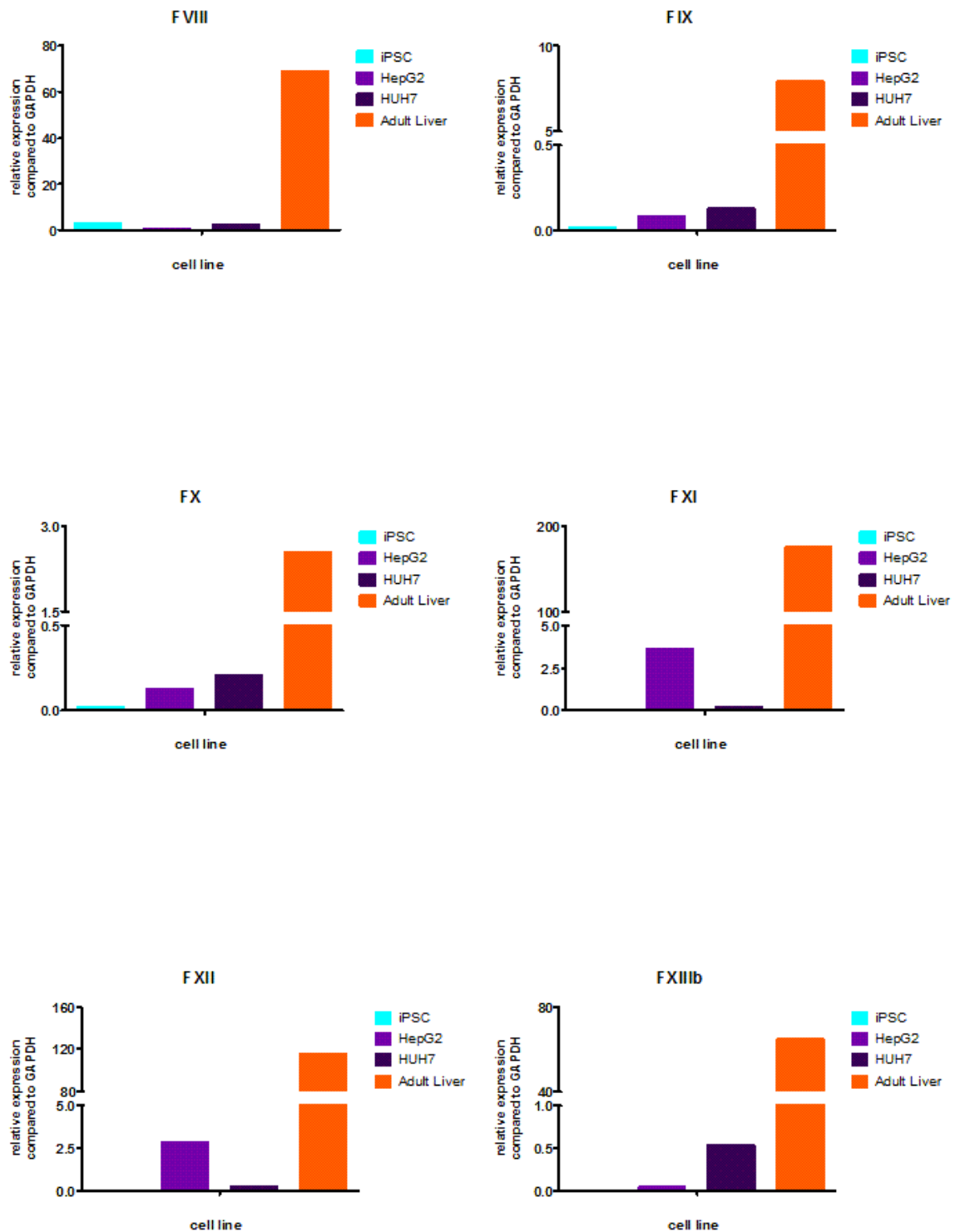


Figure 61 – coagulation gene expression in HUH7 and HepG2 cell lines. Gene expression by qPCR of coagulation genes in liver cell lines: HUH7 and HepG2. Expression levels normalized to GAPDH endogenous control.

The next step is to knockdown HNF4a and see the effect on expression of the coagulation factor genes. The use of small interfering RNA (siRNA) allows to check the effect of HNF4a knockdown in a fast and reliable way. siRNA works by causing the target mRNA to be broken down, impeding translation of the mRNA to protein. Upon introduction of the siRNA into cells, knockdown is usually achieved within 24 to 48 hours. While 48 hours are enough to assess knockdown of the target gene, we require longer periods as we want to assess the effect of the absence of HNF4a in coagulation genes expression.

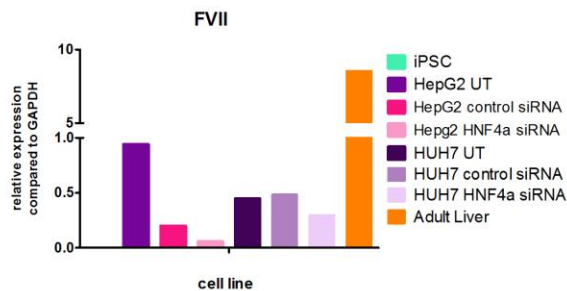
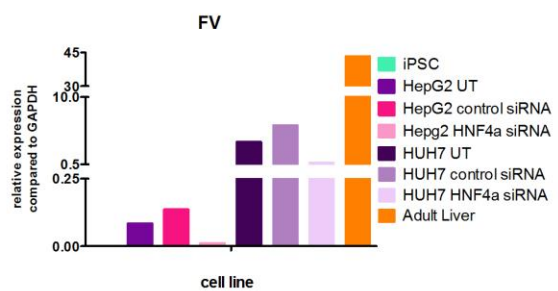
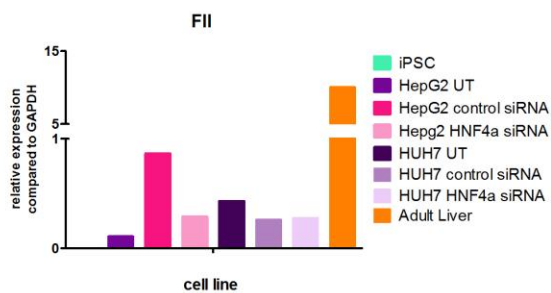
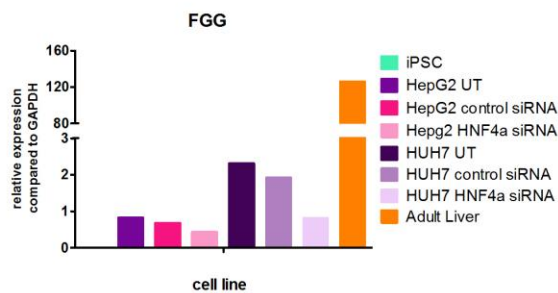
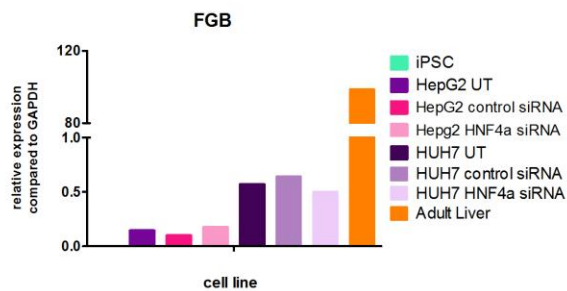
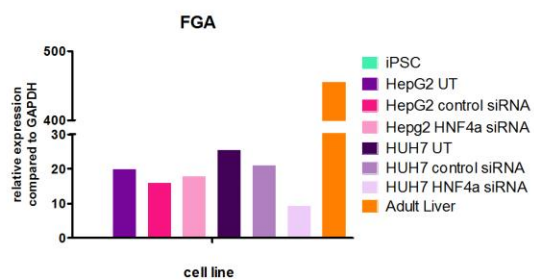
As assessed by western blot, the levels of HNF4a protein were greatly reduced in HepG2 cells in the presence of HNF4a siRNA (Figure 62). While this is promising, some protein can still be detected. Surprisingly, we also observe reduced levels of HNF4a protein in our control siRNA, which should have not happened. There is no logical explanation for this to have happened and most likely it results from human error during handling (e.g. mislabeling). Together with the absence of protein detection in HUH7 samples, it is necessary for the experiment to be repeated.



Figure 62 – HNF4a Protein levels after siRNA. Western blot of HepG2 and HUH7 cellular extracts for assessment of knockdown level upon siRNA transfection. A. HNF4a antibody B. GAPDH antibody. Legend: 1- HepG2 Untransduced, 2 – HepG2 siRNA control, 3 – HepG2 siRNA HNF4a, 4- HUH7 Untransduced, 5 – HUH7 siRNA control, 6 – HUH7 siRNA HNF4a

Regardless of the unexpected results with HNF4a knockdown with our control siRNA, we went ahead and checked for coagulation factors expression levels. If HNF4a is indeed a master regulator of the coagulation network we should see an effect on the expression levels

of the coagulation factor genes. But despite the low levels of HNF4a protein present, expression of all coagulation genes could be detected (Figure 63). This could mean that HNF4a is not a master regulator for this network. While some reduced expression can be seen in *FXI* and *FXII*, for statistical significance we need to repeat the experiment in order to increase the significance level.



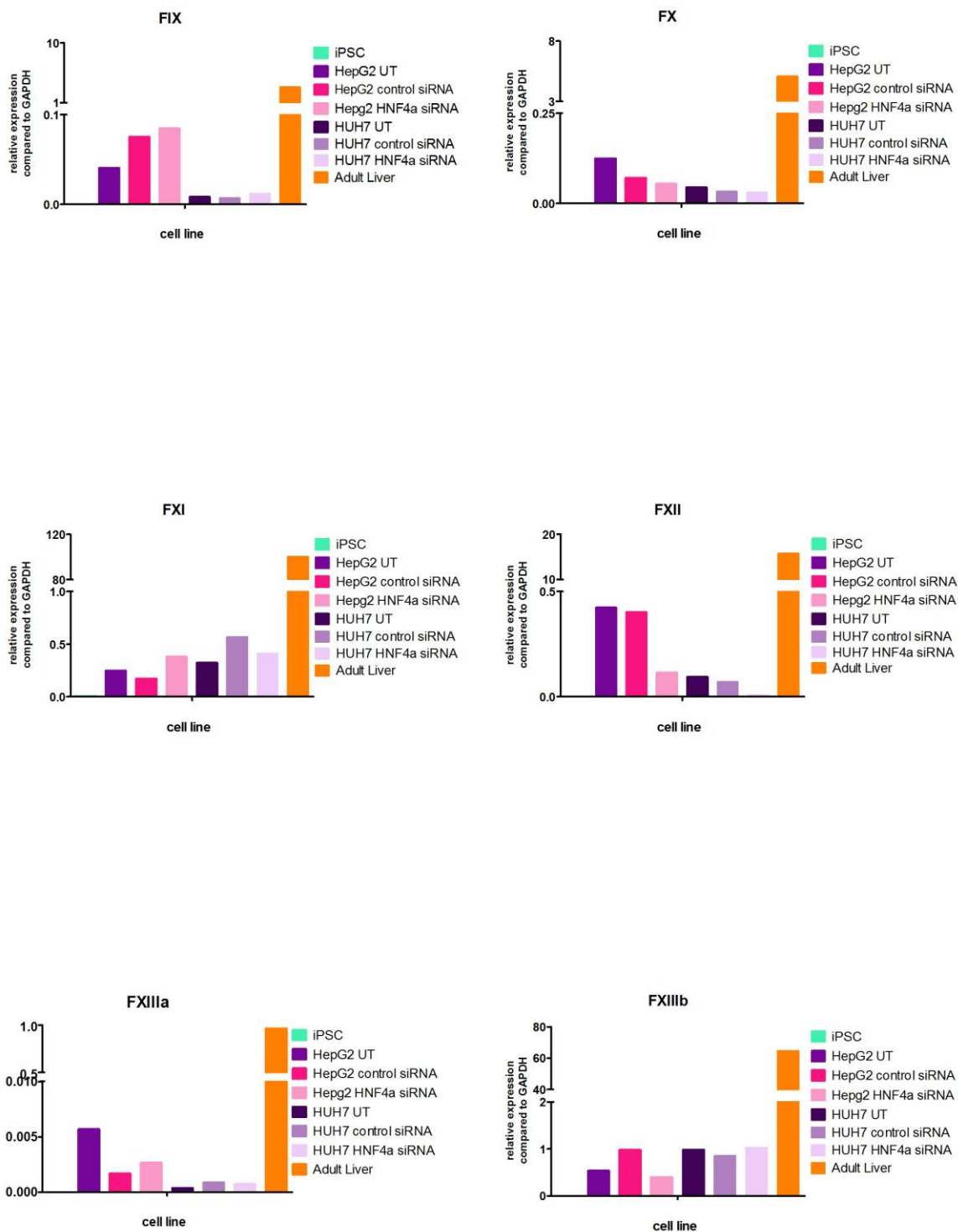


Figure 63 – Coagulation genes expression in HNF4a knockdown. Expression of coagulation genes in HepG2 cell line under 3 different conditions: untransduced (UT), control siRNA pool and HNF4a siRNA. Liver as positive control for expression and iPSC as negative control. Expression levels normalized to GAPDH endogenous control.

Considering some of the problems during siRNA knockdown, to conclusively exclude HNF4a we need to repeat the experiment. It is also important to consider other possibilities for the lack of effect from HNF4a knockdown. It is possible that the low levels of the HNF4a protein present in the cells are still enough for regulation of the coagulation genes. Another possibility can be related to the incubation period of cells with the siRNA. Even if HNF4a protein levels are low enough, the incubation period might not be enough for the knockdown of HNF4a to affect mRNA levels of the coagulation genes. It is necessary to allow time for the mRNAs to degrade while HNF4a is absent, permitting to see the real effect of HNF4a knockdown. And for that, a range of longer incubation periods need to be used.

The use of other silencing techniques, such as shRNA, would represent another alternative as these give a more sustained knock down of HNF4a, and allow to better assess its role in regulating coagulation genes network. Despite its advantages at a more stable knockdown, shRNA has the disadvantage of longer optimization periods and, it being an integrative method, can affect genomic stability.

Another important consideration relates to the list of candidate master regulators. We picked HNF4a under the belief that from the 5 candidates, it was the one with the strongest biological evidence on its role in coagulation. We also assumed that there was one master regulator for this network when it is possible that it possesses more than one master regulator. And so, to truly dissect the network and its master regulator(s), one must consider the possibility of multiple TFs taking up that role together. A good starting point would be to use the other candidates from our analysis (Table 10) as these show strong evidence as regulators of coagulation genes.

7.4 Discussion

Using the coagulation system as a model system we have attempted to use the iPS-hepatic differentiation protocol to understand the ontology of coagulation factor expression. We show that the extrinsic and common pathway coagulation factor genes are switched on relatively early during this differentiation (between days 4-8). We then tried to identify the regulatory pathway that promotes the upregulation of these genes during this time point and identified 5 possible candidates through a combination of RNA-seq and motif analysis. We hypothesised that HNF4 was likely a lead candidate because of the strong evidence connecting it to the coagulation factors. However, preliminary attempts at knockdown of HNF4 did not work as planned because of technical problems but there was a suggestion that expression of some of the extrinsic and common coagulation factor gene mRNA was modestly down regulated. These experiments need to be repeated with better siRNA reagents.

Further work needs to be done before we can definitely say HNF4a is not the master regulator of the coagulation genes regulatory network. As discussed above, a range of experiments can be done to assess this: from repeating the siRNA, using shRNA and even consider other TFs as (co-)master regulators of the network.

Even with the unanswered question, it is clear that combination of RNA-seq from our differentiation protocol samples with *in silico* methods allowed to increase the knowledge on the overall regulatory network for the coagulation genes. Despite the benefits of combining these analyses, one must be cautious with the limitations these methods possess.

While RNA-seq represented an improvement on gene expression quantification methods, one must consider how well it captures the biologic reality. RNA-seq focuses on mRNA levels, used as proxy for protein levels estimates. While a lot of information can be acquired from mRNA levels (sequence information, relative gene expression and even information on RNA splice events³⁰⁶), it is the final protein products that have the functional role we are interested in. It is important to keep in mind that a gene's expression level alone is not predictive of

protein abundance³¹⁴. Furthermore, gene expression is tissue specific³¹⁵ and even within tissues it is cell-type specific. Caution needs to be taken when interpreting data from a heterogeneous mix of cells as often it affects the data interpreted. Our model has shown to result in a heterogeneous population (Chapter 5 section 5.2.2), and while this can affect the data obtained from RNA-seq, we believe most of the variability observed results from different developmental stages of the cells rather than distinct cellular types. Presence of cells at different developmental stages can still affect data interpretation, especially as we are looking at differential expression between the different time points.

RNA-seq has proved itself extensively in the amount and quality of data it provides, and while it has been almost a decade since the first RNA-seq was performed, data collection and statistical analysis are continuously under development. Differences in the methods used for analysis can greatly affect the results, and as large amounts of data are obtained from single experiments, it is possible for important data to be overlooked. These are true for any RNA-seq data presented, and is a general limitation rather than specific to our experiment.

Data from *in silico* analysis are predictive of biological events based on theoretical knowledge from previous research. A variety of algorithms has been developed for different analysis, such as motif analysis, with most showing low levels of predictive accuracy³¹⁶. Increased accuracy can be obtained when biological knowledge is incorporated in the analysis, such as high-throughput data. Accounting for chromatin modification is an important example of biological data as it distinguish between TFBS that are accessible or not, and therefore their potential binding capacity under a given cellular conditions³¹⁷.

As this was done for coagulation, other liver specific pathways can be examined using this RNA-seq data in combination with *in silico* analysis. This avenue of research can also be used for other cell types, exemplifying how iPSC research can be used in ways normally overlooked. Most of the work with iPSCs, and stem cells in general, are greatly focused in 3 main fields: disease modelling, drug toxicity and cellular therapies.

The stem cell field has great potential for research and therapy and as the field evolves new possible applications have come to light. While some hurdles still exist, the benefits are clear with a promising future for the field of iPSC.

Chapter 8 Final Discussion

A great deal of work has been done regarding iPSC, both at the reprogramming and direct differentiation level.

iPSC reprogramming has dramatically changed since it was first reported in 2006, with a wide range of methods currently available. All the different factors that can influence the reprogramming process, led to an unintentional barrier against their clinical use. The multitude of ways iPSC can be obtained results in populations that differ from each other. How their differences affect their quality is still being investigated, and until a standard criterion on the quality of these cells is set, cautious is recommended. Considerations in terms of transcriptional signatures, epigenetic status, genomic integrity, stability, differentiation and tumour potential should be the focus of quality control of iPSC.

While a standard iPSC has not been defined yet, some methods have known detrimental effects on the quality of the final population, as is the case with integrative methods. As these cells can be used in different contexts, it has been proposed that different evaluation criteria could be employed³¹⁸. Moreover, new avenues are emerging by which iPSCs without the ability to form teratomas might actually be more useful and safer for regenerative medicine. While much speculation exists on which are the best criteria to evaluate iPSC quality, a consensus on this would allow to remove variability when comparing results from different labs.

With this in mind, we established a robust method of reprogramming that fulfils the current safety guidelines. The use of an integration-free, xeno-free method allows to remove most of the components that introduce variability to the process (feeders and serum containing medium) while maintaining genetic stability of the final population. These features are important in the context of clinical translational but also in establishing guidelines that can be followed for standardization of the reprogramming protocol. Despite the advantages of our adopted method, the variety of methods available has taught us that there is no one way when

it comes to reprogramming, and that similar quality iPSC can be obtained from different protocols.

The adoption of a standard quality iPSC would also level the field of direct differentiation. Forward differentiation from PSCs has been observed to almost every cell type of the human body, with varying degrees of success. For one cell type, several differentiation protocols exist and the use of similar or identical starting populations would allow direct comparison in their efficiency/ability to do so. As with reprogramming, much has improved on the quality of differentiation protocols, but further improvements are still required. These try to recapitulate the *in vivo* developmental cues in an artificial setting and as with most artificial settings, they lack all the cues provided but the cells' natural niche.

It is accepted that differentiated cells most closely match embryonic or foetal stages of development rather than adult. The physical and chemical properties of the cellular environment have recently earned interest as a way to improve the developmental state of cells. The cell-material interface is not static with cooperative interactions dictating one another's fate³¹⁹. Differentiation can be affected by the substrate mechanical stiffness^{320, 321}, the topography^{322, 323} and chemical composition^{324, 325}.

Composition of the material coupled with the concept of degradation by-product was realised to influence stem cell fate. An example is the release of ions (such as calcium, magnesium and fluoride) from dissolving inorganic mineral and its effects on stem cell phenotype^{326, 327}. This concept can be applied to a range of natural and synthetic materials. Interestingly, this effect can also influence cell shape³²⁸ which also leads to changes in differentiation.

Topography cues of the environment, such as shape can affect cell phenotype. Shape affects cytoskeletal organization of cells which in turn has shown to affect cell lineage fate³²⁹ in a "function follows shape" manner. A known example of topography effect is on cardiomyocyte functionality and maturation. Cardiomyocytes are muscle cells that possess contractibility features. While *in vitro* differentiation protocols result in cells that show a beating phenotype, these cells lack correct morphology, structure organization, mechanical output and electrophysiology. Culture of cells in rectangular patterns allowed cells to obtain

their typical elongated morphology, resulting in improved phenotype which is closer to mature cells³²³. Improvements in structural and electrophysiological features of cardiomyocytes show us how topography of the substrate can contribute to the maturation of the final population.

The same way composition and topography can affect cell fate, stiffness of the substrate can elicit different responses in cell maintenance or differentiation. This has been shown for a range of different substrates, including collagen³³⁰, polymer networks³³¹, PDMS³²¹ and even nanofibers³³². Stiffness can induce different cellular responses as it affects the signals provided from the interactions with the substrate, which in turn can modulate internal organization of cells.

Although these mechanisms have been discussed in their individual ability to regulate cell fate, they work together in complex ways with the process being more challenging than just linear combination of the different inputs. Indeed, similarly to what happens in biochemical induced differentiation, material-induced differentiation might require multiple steps by which the appropriate signals are provided in a specific temporal sequence for correct differentiation.

This view that cell fate can be influenced by composition and physical properties of the substrate clashes with the traditional view by which molecules (such as growth factors) are the main drivers of differentiation. The reality is that the complex developmental cues cannot be recapitulated by individually using material or media, but rather a combination of all the possible technologies presents the best chance at obtaining optimal differentiation protocols.

Along this train of thought, and with the realization of the limitations from 2 dimensional (2D) culture systems in replicating the *in vivo* environment, 3D systems started being employed. The differentiation process in 2D can have varying levels of efficiency, with the worst cases showing efficiencies <10% as is the case for pancreatic endocrine cells³³³ while neuron differentiation can achieve >80% efficiencies¹¹⁸. Even if efficiencies are satisfying, quality of the differentiated population is an important consideration. Together with the physical and chemical properties of the environment, 3D structures try to bring the

environment closer to the cell's natural niche. 2D systems are known to reproduce inadequately the *in vivo* features, as exemplified by the interaction of cells with the ECM and neighboring cells via the basal region, affecting protein cell-surface distribution and the organization of the intracellular machinery and thus changing cellular responses^{334, 335}. Differentiation in 3D scaffolds has showed to improve the quality of the differentiation³³⁶, organization³³⁷ and functionality^{337, 338} of cells, and different lineages have been obtained including neuronal³³⁹, cartilage^{337, 340}, vascular^{336, 341} and liver³³⁸. In some cases the combination of 2D and 3D systems has proven advantageous, as in the case of intestinal tissue, with an initial stage in 2D for hindgut differentiation followed by a 3D step that allows polarization into villus-like structures and crypt-like proliferative zones that express intestinal stem cell markers³⁴².

The growing movement in 3D culture together with research in developmental biology and cell dissociation and re-aggregation allowed for organoids to be derived from human PSCs, including gut, kidney, brain, retina and liver structures³⁴²⁻³⁴⁵. Organoid is defined as an organized structure of organ-specific cell types that develops from stem cells or organ progenitors and that exhibits some organ specific function. As with the case of PSCs, organoids have great potential in regenerative medicine, including disease modelling, drug toxicity and cell or whole organ replacement therapy.

Organoid development required a great deal of optimization as several factors can affect the success of a 3 dimensional structure forming: medium composition, density of cells, developmental stage of cells, temporal cues, among others. In some cases, success in organoid formation relied on presence of extracellular matrix components, such as matrigel for neural organoid formation³⁴³. However, the requirements of a 3D organization are variable. Some studies show that structural organization of cells is not a requirement for their functional use^{346, 347}.

A situation where a 3 dimensional structure has proven beneficial for functionality of differentiated cells was the development of liver-bud structures. This was done by mixing three different populations: human PSCs derived hepatic cells, human mesenchymal stem cells and human endothelial cells, as way of mimicking the early cell lineages of the

developing liver³⁴⁵. When mixed, the cells spontaneously aggregate into a 3D structure that resembles a liver bud. These organoid structures presented vascularization and when transplanted into mice showed ability to rescue liver failure and production of human specific metabolites³⁴⁵. Recreation of liver architecture allows for correct cell-cell and cell-matrix interactions to be formed and hepatocyte polarity to occur, which has been shown to be important for correct liver function³⁴⁸.

There are two known mechanisms known to influence organ self-assembly: the first, known as Steinberg's differential adhesion hypothesis, says that cells with similar adhesive properties segregate into domains that achieve the most thermodynamically stable pattern³⁴⁹; the second being "proper spatially restricted progenitor fate decisions", by which stem cell division, and orientation, balance of symmetric and asymmetric divisions, and migration of differentiated cells are required to occur in a correct way for normal tissue architecture development. This demonstrates the importance of the interplay of the different cell types that make up a tissue for proper differentiation of cells.

Organoid technology carries its own limitations such as vascularization of the structures. Most organoids do not possess vasculature which imposes limitation in nutrient supply. This can affect the growth potential, and while spinning conditions have shown to improve nutrient exchange, the use of a function vascular network has proven the best alternative. Co-culture with endothelial cells has shown to generate a vascular-like network with transplantation into recipient mice showing incorporation of this network into the host vasculature³⁴⁵.

Furthermore, organoids resemble more primitive developmental structures and for full maturation, transplantation into a recipient might be necessary. The maturation state of cells is a recurrent issue in stem cell derived populations. It is possible that until better methods are developed, an *in vivo* step is the only way maturation of cells can be achieved. This shows that despite how far technologies have come in improving differentiation from stem cells, there are still hurdles along the way to be overcome.

The use of iPSC for cellular therapy depends on robust differentiation protocols that are able to produce cellular products able to rescue the disease or damaged phenotype. While maturation level of cells is still a concern, the use of PSCs as starting population remains a safety issue for many. A lot is still unknown regarding the quality of iPSC when it comes to genetic stability, and while resetting the epigenetic state of cells is a key requirement for the reprogramming process, it incurs risks for the quality of the final population.

The conceptual idea of transdifferentiation (Figure 2) becomes highly relevant for cell fate conversion without requirement of passing by a pluripotent state. By the introduction of specific transcription factors, direct conversion of somatic cells has been observed to a range of different lines^{350, 351} both intralinear and across lineages (Table 11). By avoiding a pluripotent state, we might secure a higher genetic stability of cells. Overall, it is easier to transdifferentiate cells into related lineage phenotypes as their epigenetic signatures are closer and therefore easier to convert into one another.

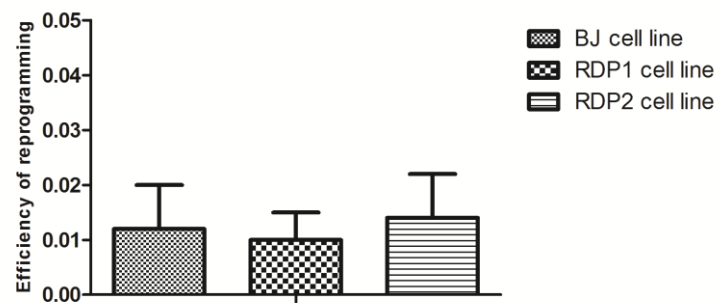
Table 11 – Transdifferentiation between somatic cell states

Somatic cells conversion (with germ layer)	Exogenous reprogramming factors	Experimental setting	Reference
Fibroblasts (mesoderm) converted to myocyte-like cells (mesoderm)	MyoD + 5-Azac	<i>Invitro</i> Intralineage conversion	352
B cells, T cells and Fibroblasts (mesoderm) converted to macrophage-like cells (mesoderm).	C/EBP ± PU.1	<i>Invitro</i> Intralineage conversion	353
Cardiac fibroblasts (mesoderm) converted to induced cardiac myocyte-like cells (mesoderm)	Gata4, Mef2c and Tbx5	<i>Invitro</i> Intralineage conversion	354
Fibroblasts (mesoderm) concerted to induced neuron- like cells (ectoderm)	Ascl1, Brn2 and Mytl1	<i>Invitro</i> Cross-lineage conversion	355
Pancreas exocrine cells (endoderm) converted to endocrine-like cells (endoderm)	Ngn3, Pdx1 and MafA	<i>Invitro</i> Intralineage conversion	356

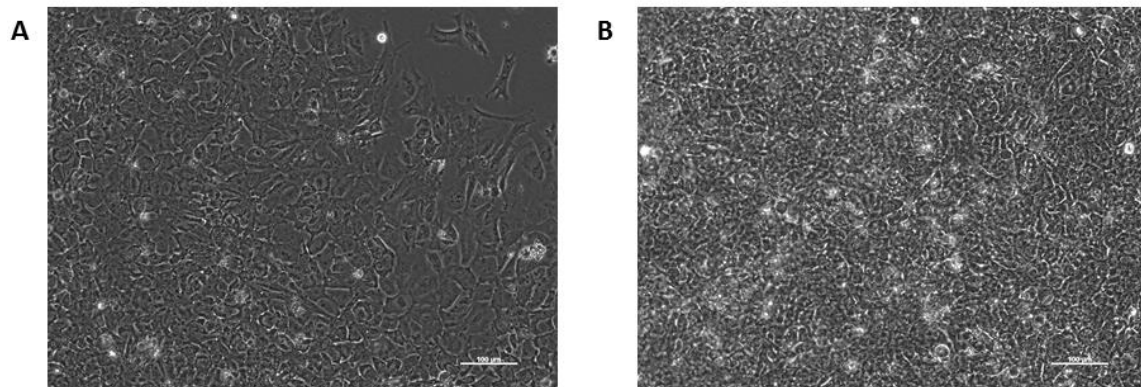
The field of stem cell has seen an exponential level of development in the past years, with progresses in other areas permitting such growth. A great deal more is expected in the following years and hopefully the potential of this technology will soon be fulfilled.

Chapter 9 Supplementary Information

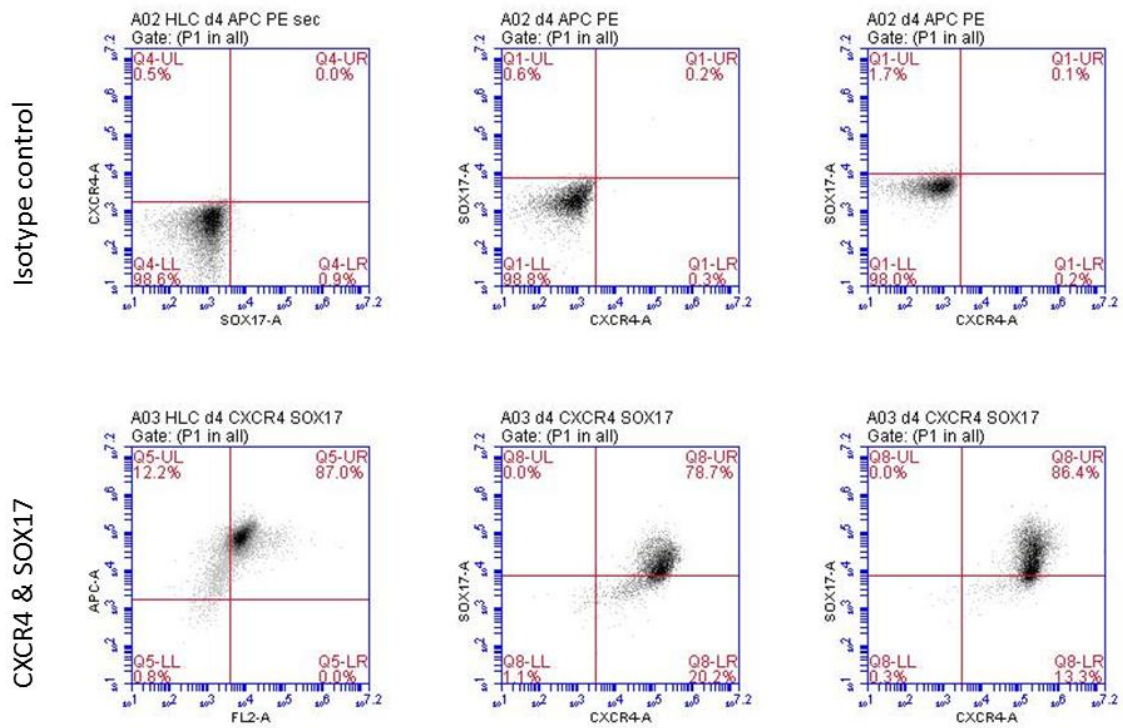
9.1 Supplementary Figures



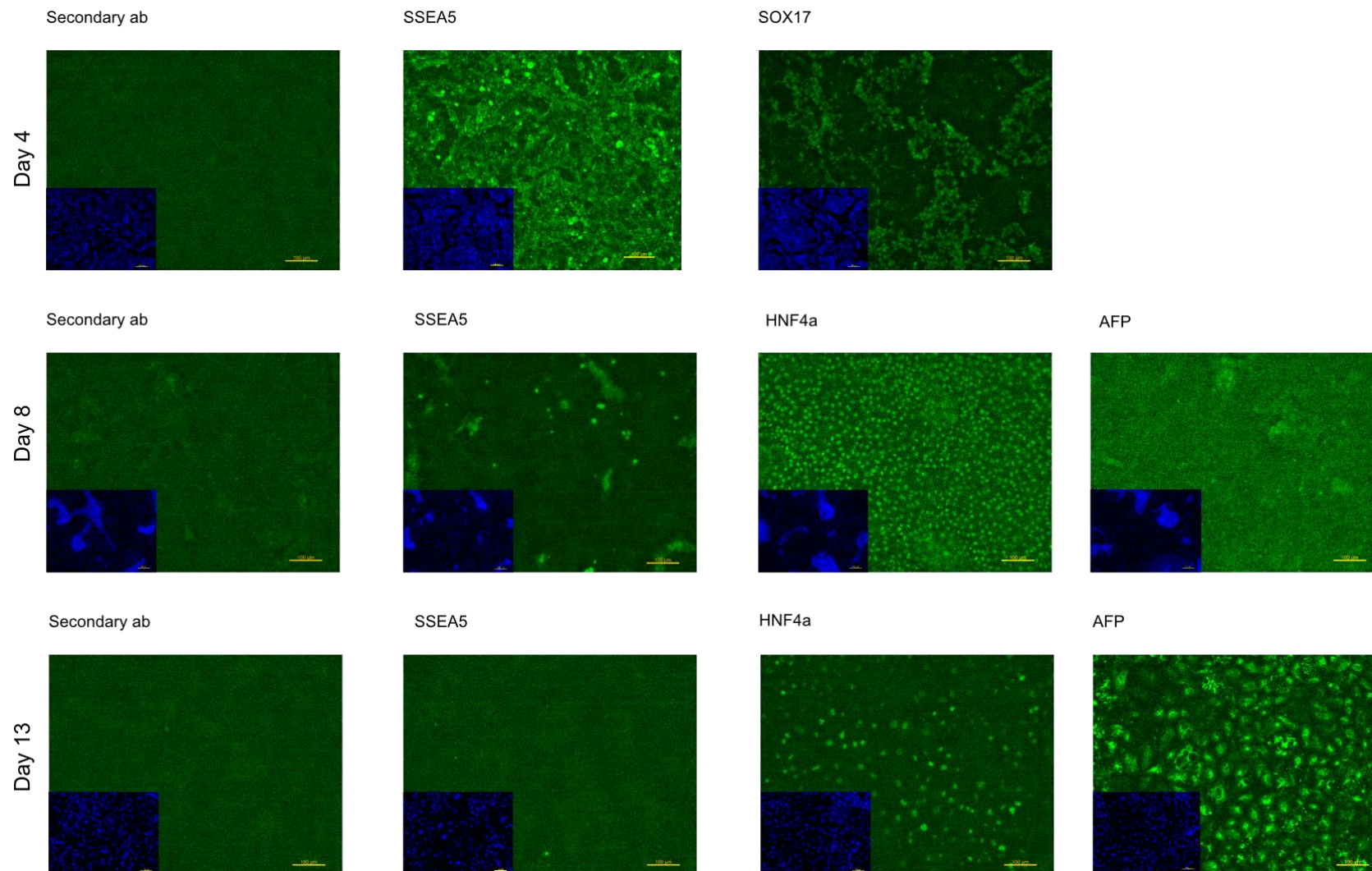
Supplementary figure 1 – Reprogramming efficiency of different cell lines using episomal method. All cell lines showed similar reprogramming efficiencies using the episomal method. In brief, 10^6 human fibroblasts were transfected with Yamanaka's episomal plasmids (1 μ g each) using Amaxa Nucleofector kit and cultured in MEF medium and Nuff feeders.



Supplementary figure 2 – Endoderm differentiation in iPSC from colony format. Endoderm differentiation from colony format results in different condition for cells at the periphery and centre of the colony A. cells at periphery of the colony. B. cells at the centre of colony

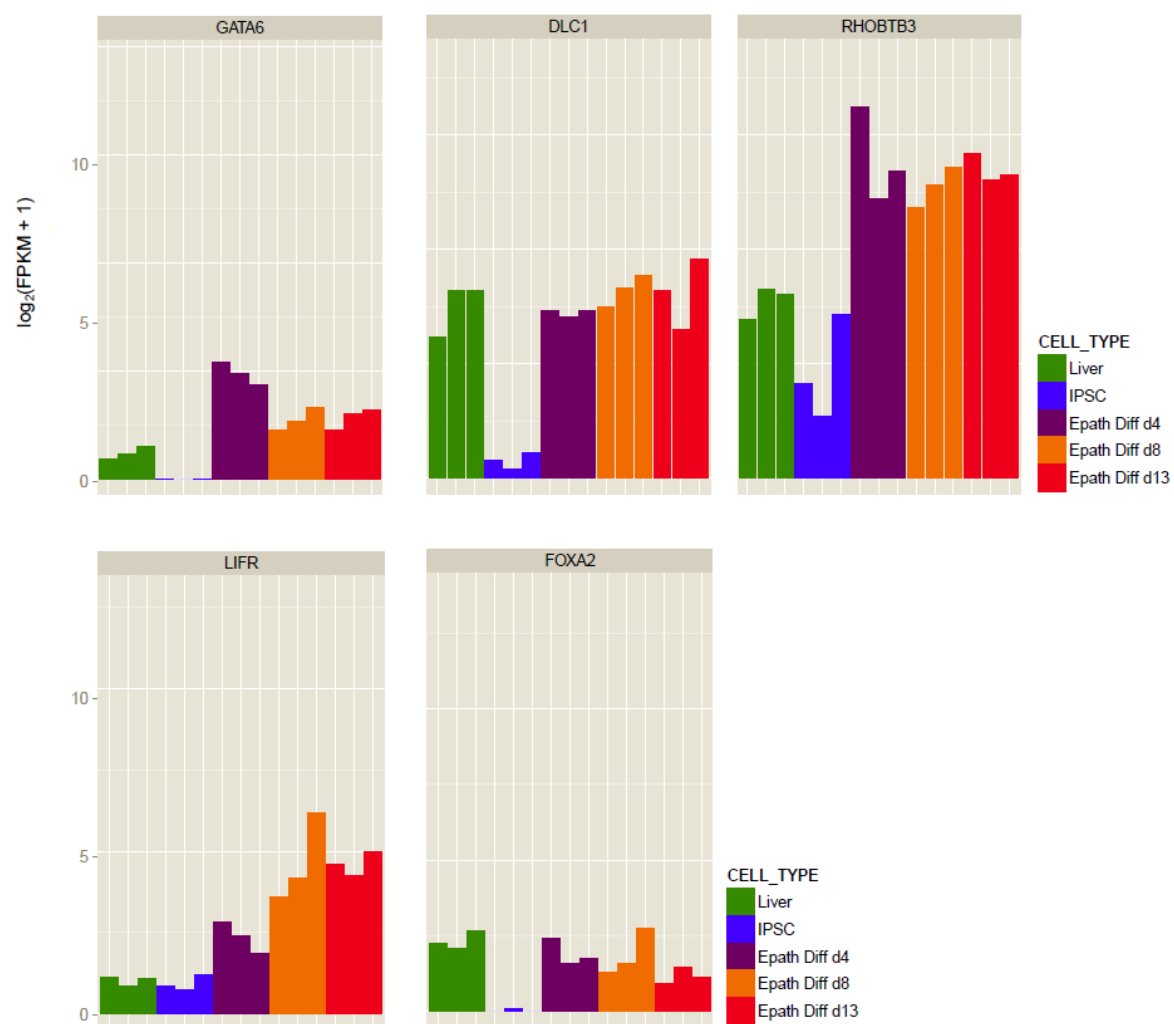


Supplementary figure 3 - FACS characterization of HLC day 4 using endodermal specific markers CXCR4 and SOX17. . FACS analysis of 3 different cells lines at day 4 of the hepatic differentiation protocol. Endodermal cells from different iPSC lines all show high levels of expression of endoderm surface markers: CXCR4 and SOX17.

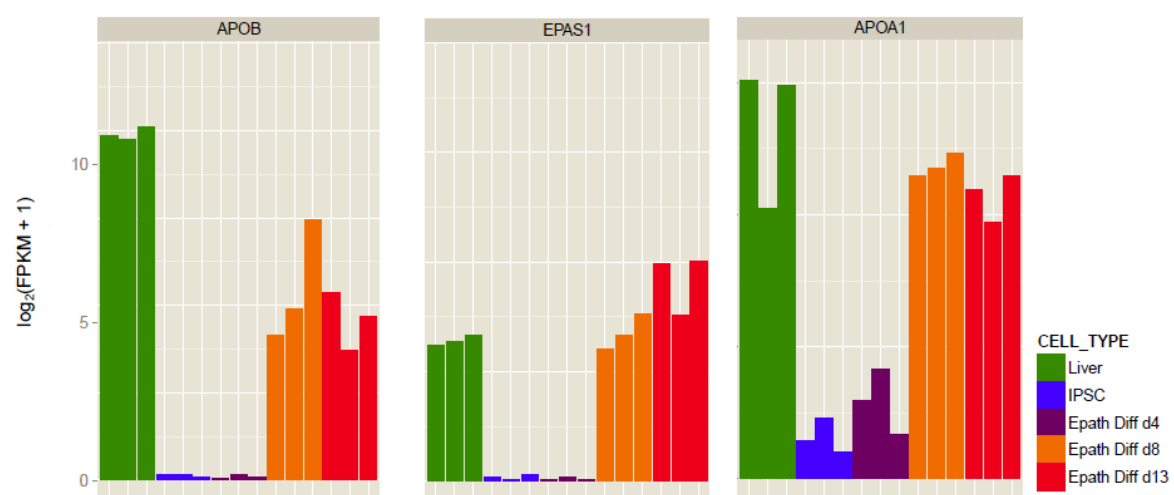


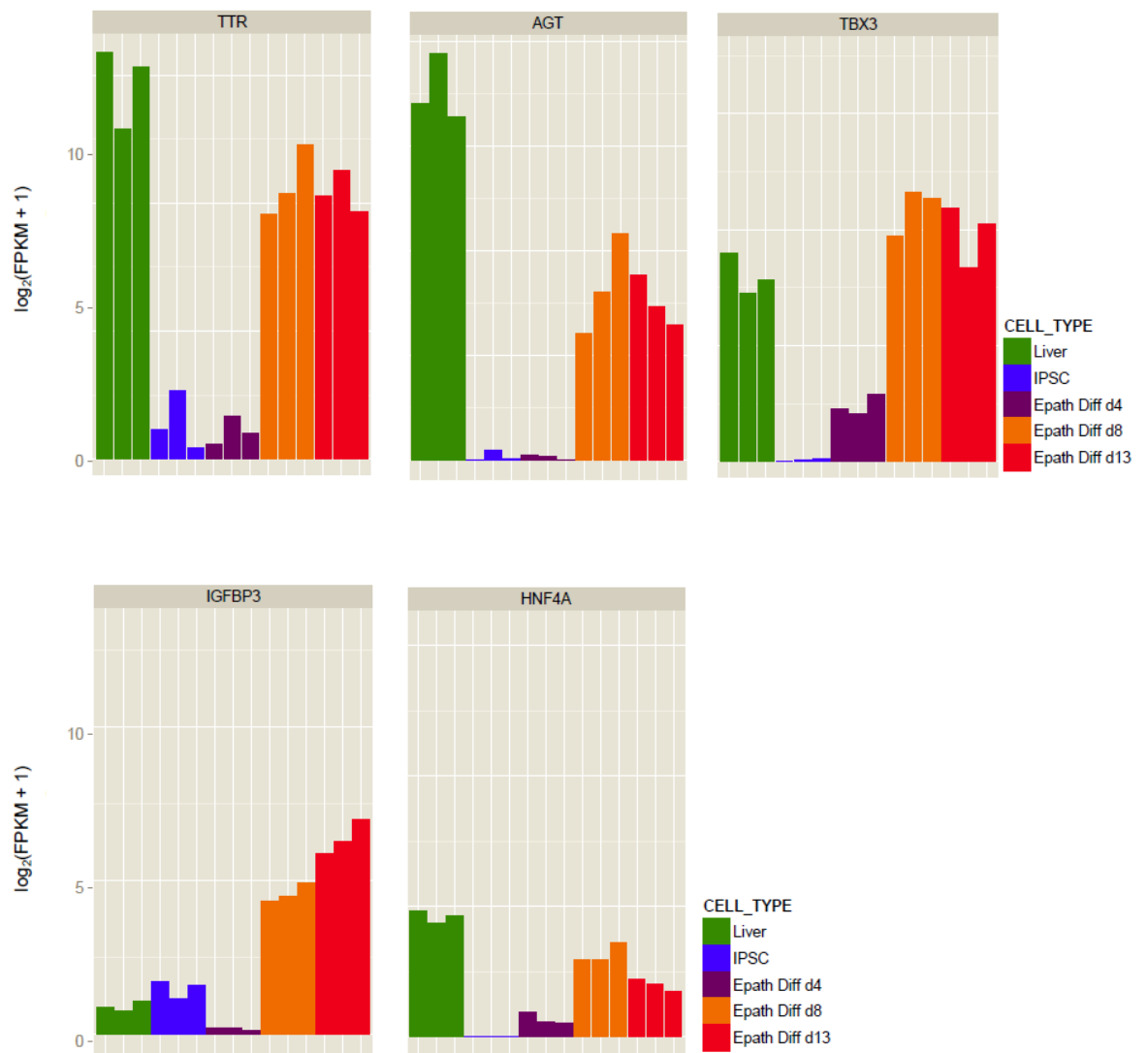
Supplementary figure 4 – Immunostaining of cells during hepatic differentiation protocol. Analysis using a range of different markers to characterize cells during hepatic differentiation protocol: SSEA5 (pluripotent marker), SOX 17 (endodermal marker), Nestin (ectodermal marker), HNF4a and AFP (Liver specific marker).

A

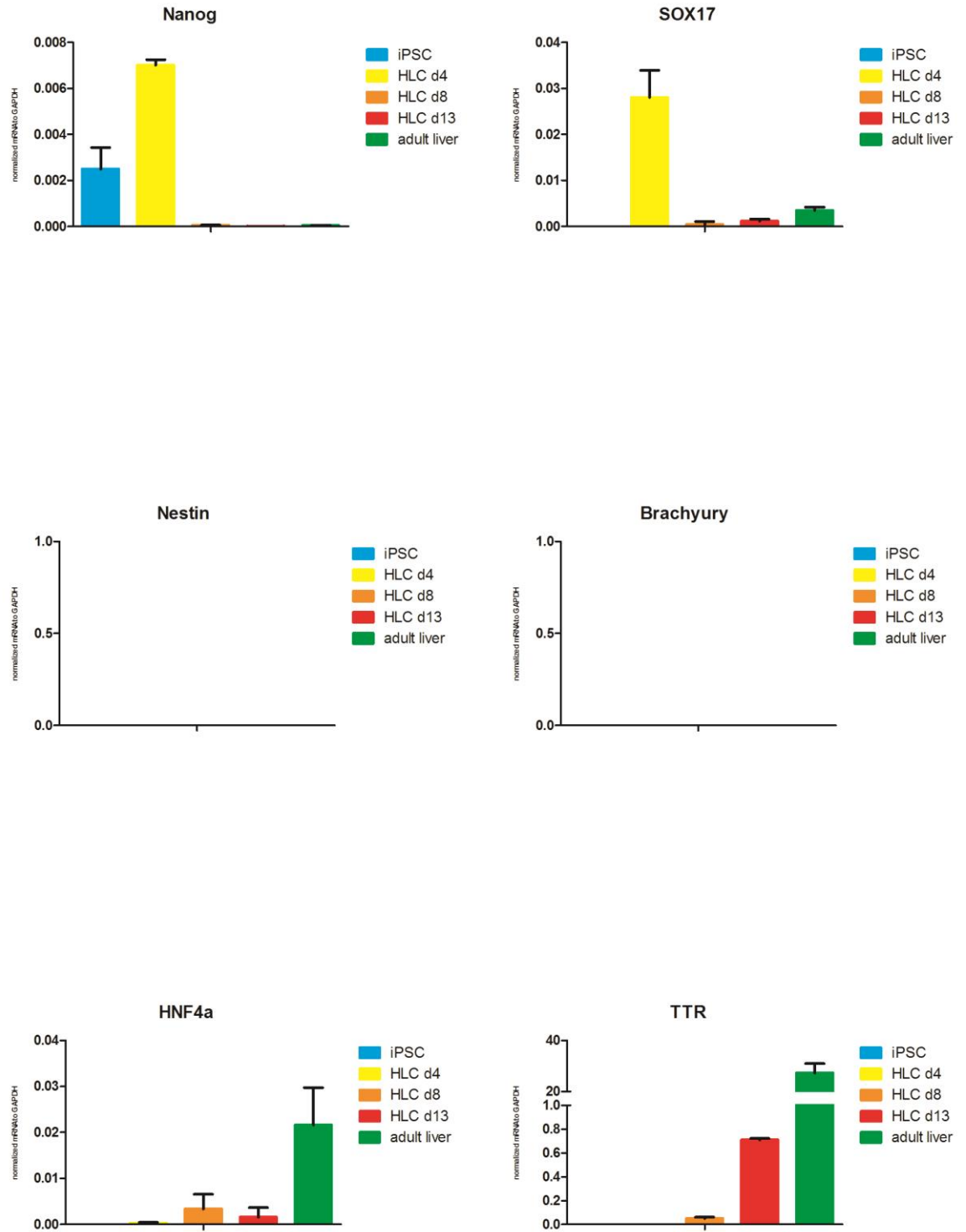


B

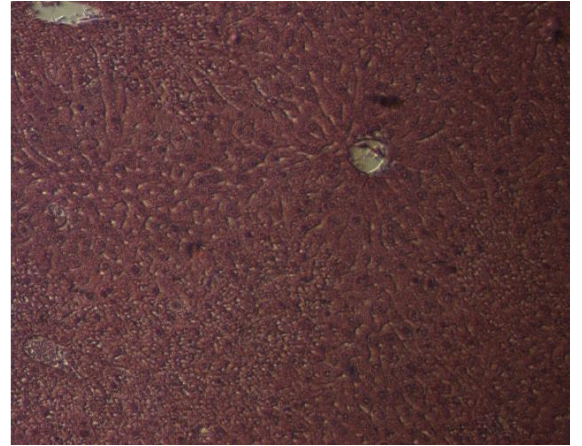
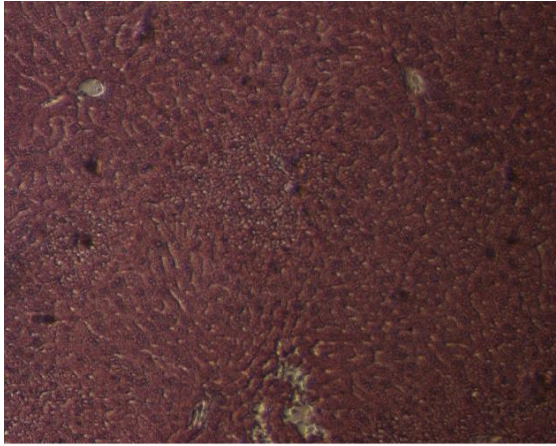




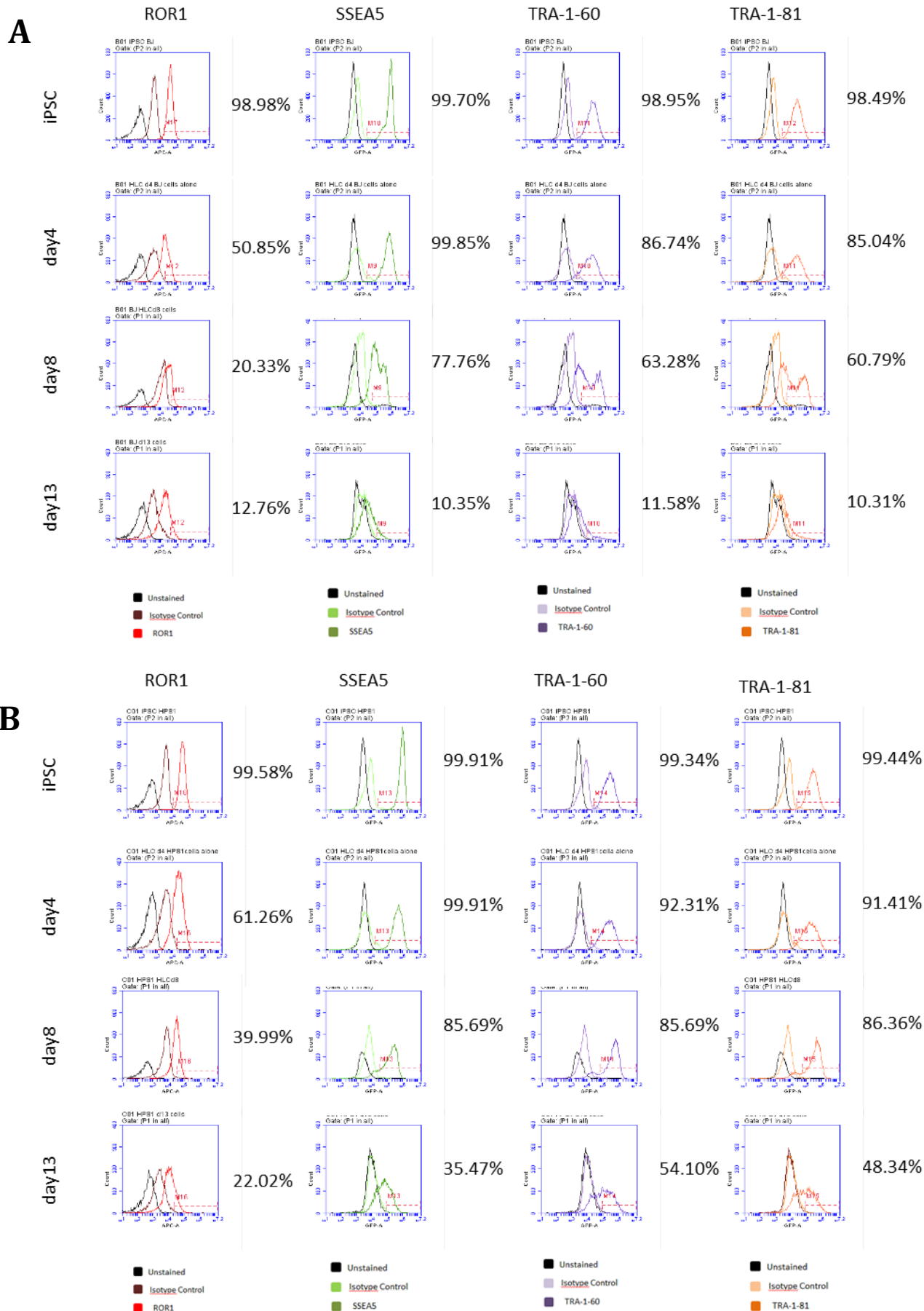
Supplementary figure 5 - RNA sequencing analysis. A. Genes turned on at the endoderm stage and present throughout the hepatic differentiation protocol. B. Genes turned on at the hepatic commitment stage and present throughout the hepatic differentiation protocol.



Supplementary figure 6 – Gene expression of cells during hepatic differentiation protocol. Analysis using a range of different markers to characterize cells during hepatic differentiation protocol: Nanog (pluripotent marker), SOX 17 (endodermal marker), Nestin (ectodermal marker), Brachyury (mesodermal marker), HNF4a and TTR (Liver specific marker).



Supplementary figure 7 – Liver of mice treated with CCL4. H&E of mice livers after CCl₄ treatment. Normal architecture can be observed as opposed to first round of CCL₄ treatment that resulted in damaged liver.



Supplementary figure 8 - Expression of surface markers assessed by FACS. Assessment of pluripotent and tumorigenic associated surface markers in cells during hepatic differentiation A. FACS profile of BJ iPS cell line during hepatic differentiation. B. FACS profile of HPS1 iPS cell line during hepatic differentiation.



BJ iPSC



BJ HLC day 4

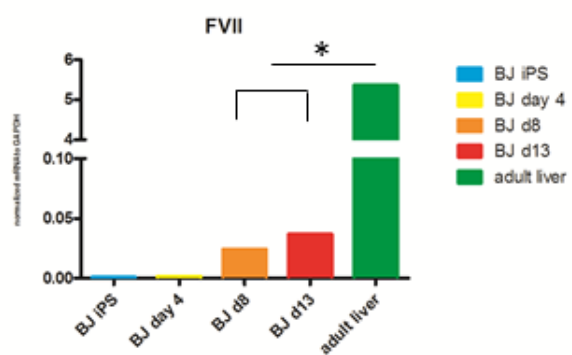
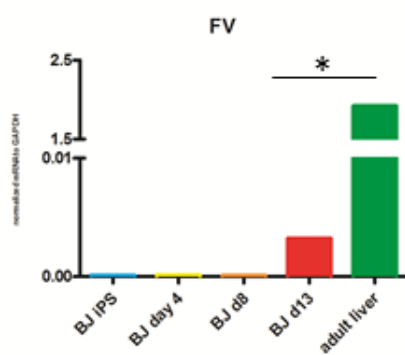
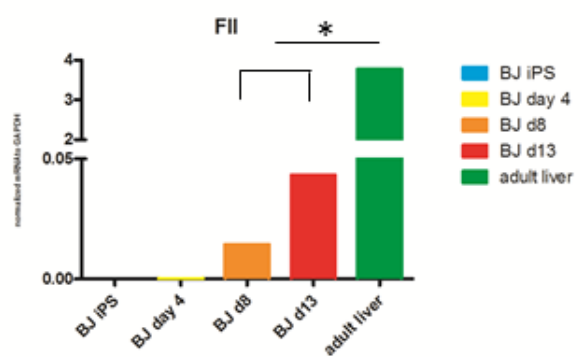
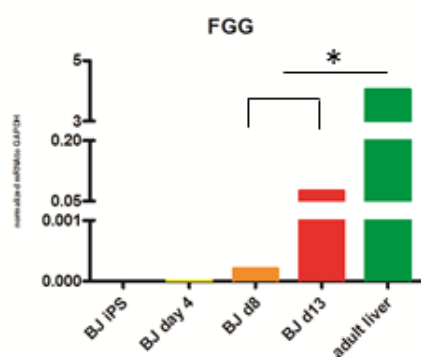
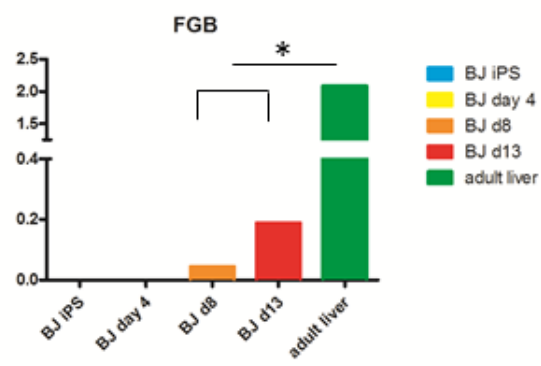
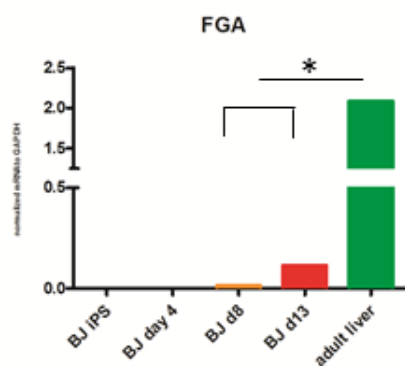


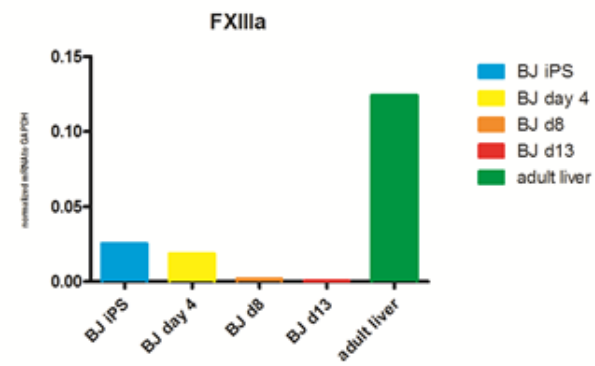
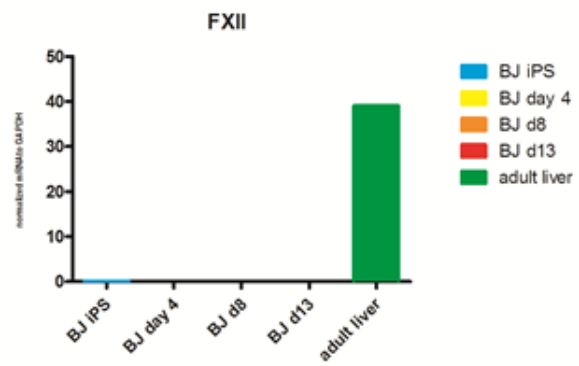
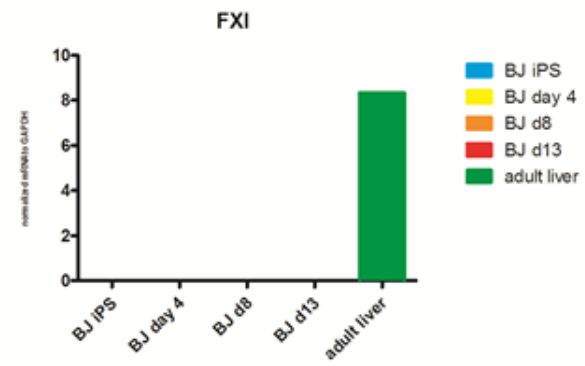
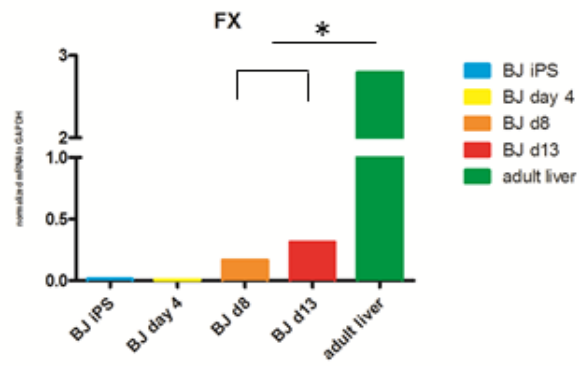
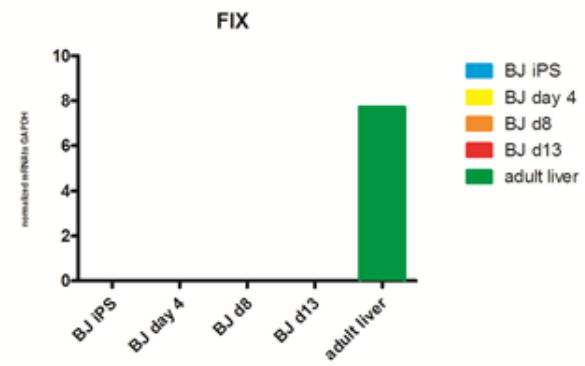
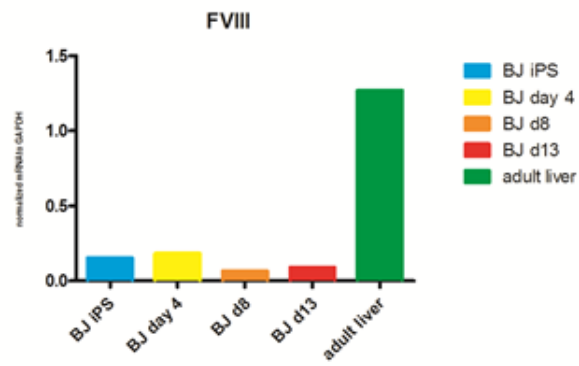
BJ HLC day 8

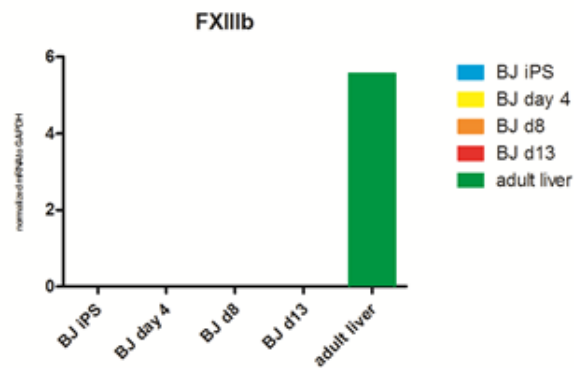


BJ HLC day 13

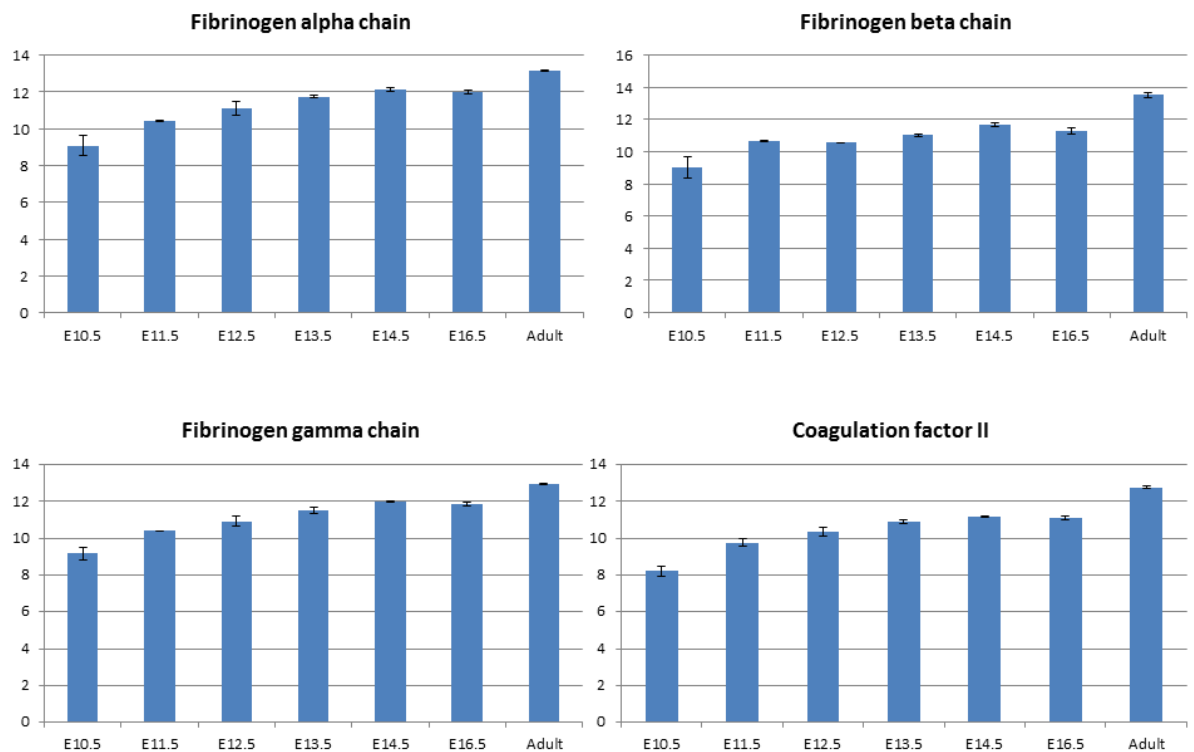
Supplementary figure 9 – Teratomas from HLCs at different time points. Macroscopic pictures of tumours after dissection. Tumours/testis are kept in Histochoice until sent to Histology.

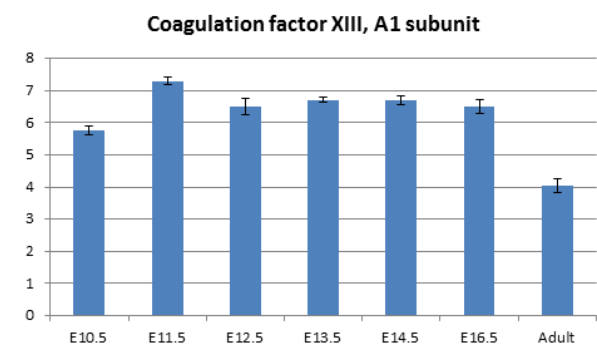
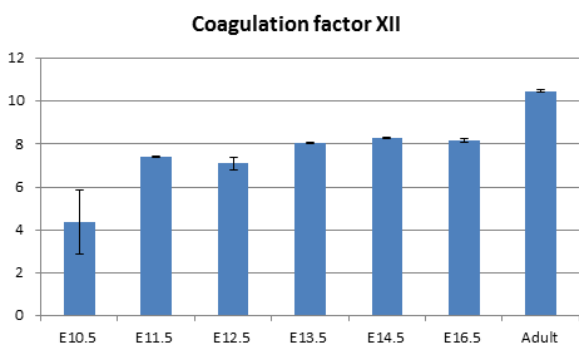
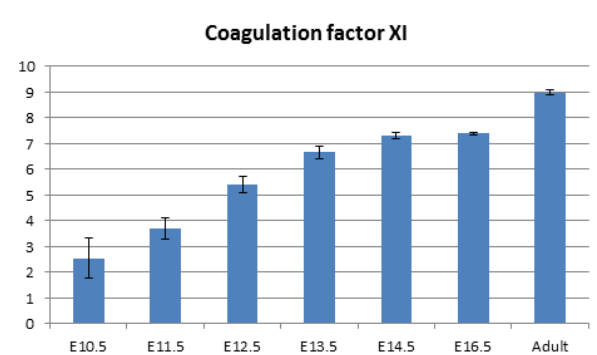
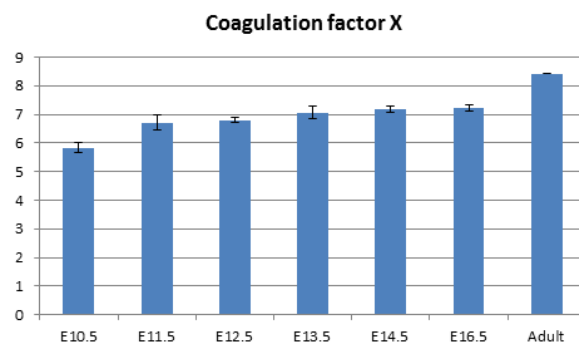
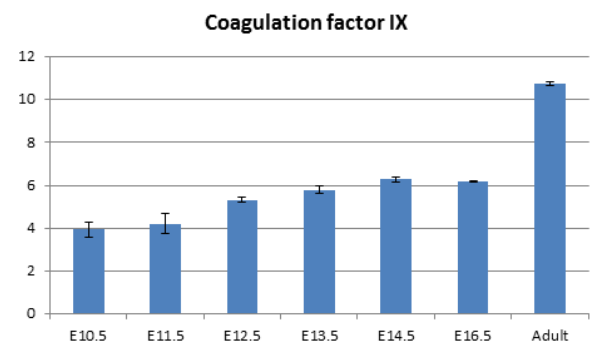
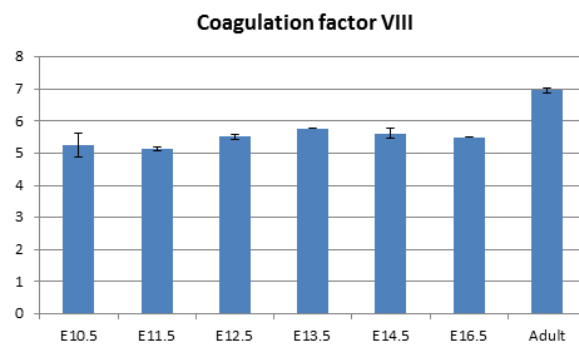
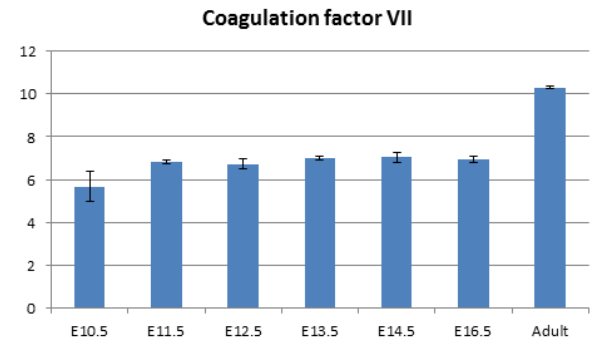
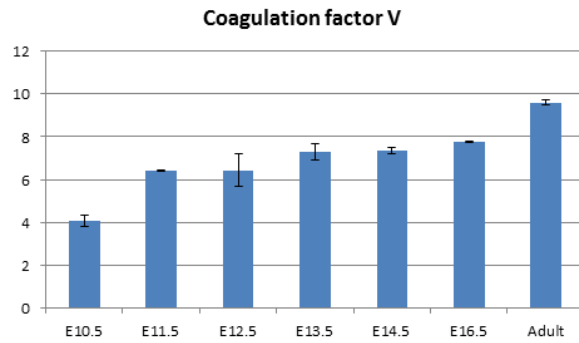


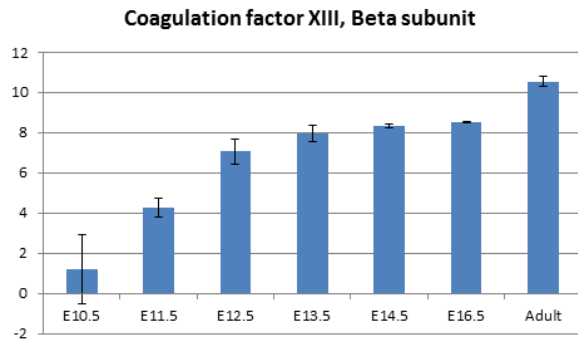




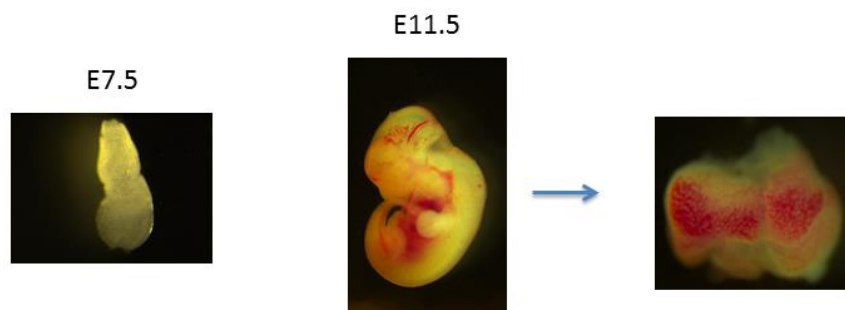
Supplementary figure 10 - Coagulation factors gene expression data in BJ iPS line hepatic differentiation. Gene expression by qPCR at different time points of hepatic differentiation protocol: day 0 (iPS), day 4, day 8 and day 13. Adult liver as positive control. Expression levels normalized to GAPDH endogenous control. Stats using unpaired t test. Legend: * statistically significant ($p < 0.05$)







Supplementary figure 11 - Coagulation factors expression in mouse embryonic liver samples in GEO dataset. Expression analyzed in Transcriptome Analysis Console (TAC) from Affymetrix



Supplementary figure 12 - Pictures of mouse embryos during dissection. Example of E7.5 and E11.5 of mouse embryos.

9.2 Supplementary Tables

Supplementary table 1 - In-house Database of protein coding genes with annotations for tissue specificity and transcription factors. The DB is built in MySQL using Relational Tables to match IDs and downloadable as .csv and .xml for analysis in R

Number Of Genes	Description	Source
23044	All Human Genes	Ensembl and Uniprot
1406	All Human Transcription Factors	Jaspar, Aftb, Transfac, Uniprot, Msig Db, Fantom Project
4721	All Human Liver Genes	Uniprot, Literature Search, Human Protein Atlas Project
1453	All Human Liver Transcription Factors	Fantom, Protein Atlas, Literature Search
867	Definite Liver Tfs	Fantom, Protein Atlas, Literature Search, Uniprot, Manual Curation

9.3 Motif analysis

A master database of ~1400 human transcription factors was built in-house by collecting data from Uniprot, TRANSFAC, Protein Atlas Project, MEME suite, Jaspar Database and data mining through biomedical publications. Matrix profile with frequencies of each nucleotide in columns were collected from Jaspar and TRANSFAC databases and results from Jolma *et al.*³⁵⁷ and Vaquerizas *et al.*³⁵⁸. The matrices are built based on the consensus nucleotide in each position having the highest score in column (Supplementary figure 13). Think of consensus as an “ancestor” motif, from which mutated motifs evolved. Matrices obtained from JASPAR, TransFAC and MSIGDB. In-house scripts to search for sequences from the matrices across the 15kb upstream sequences of the coagulation genes.

Using an in-house script for treating duplicates and element searching, we parsed the TFBS matrices across the promoter sequences to check for potential regulatory regions in the coagulation genes. According to the hypothesis of this study, master regulators are those that regulate most of the genes, in the coagulation network. Therefore, from the hundreds of transcription factors predicted to be binding to the 14 coagulation genes, only 95 possible transcription factors were common to coagulation genes and were short listed.

Alignment

a	G	g	t	a	c	T	t
C	c	A	t	a	c	g	t
a	c	g	t	T	A	g	t
a	c	g	t	C	c	A	t
C	c	g	t	a	c	g	G

Profile

A	3	0	1	0	3	1	1	0
C	2	4	0	0	1	4	0	0
G	0	1	4	0	0	0	3	1
T	0	0	0	5	1	0	1	4

Consensus

A C G T A C G T

Score

3+4+4+5+3+4+3+4=30

$$\text{Score}(\mathbf{s}, \text{DNA}) = \sum_{i=1}^l \max_{k \in \{A, T, C, G\}} \text{count}(k, i)$$

Supplementary figure 13 – Schematic representation showing how a TFBS binding at a specific sequence pattern is matched and scored

REFERENCES

1. Thomson, J.A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147 (1998).
2. Briggs, R. & King, T.J. Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Zoology* **38**, 455-463 (1952).
3. Gurdon, J.B. The developmental capacity of nuclei taken from intestinal epithelial cells of feeding tadpoles. *Journal of Embryology and Experimental Morphology* **10**, 622-640 (1962).
4. Gurdon, J.B., Laskey, R.A. & Reeves, O.R. The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs. *J. Embryol. Exp. Morph.* **34**, 93-112 (1975).
5. Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J. & Campbell, K.H.S. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810-813 (1997).
6. Kato, Y. *et al.* Eight calves cloned from somatic cells of a single adult. *Science* **282**, 2095-2098 (1998).
7. Meng, Q., Polgar, Z., Liu, J. & Dinnyes, A. Live birth of somatic cell-cloned rabbits following Trichostatin A treatment and cotransfer of parthenogenetic embryos. *Cloning and Stem cells* **11**, 203-208 (2009).
8. Wakayama, T., Perry, A.C.F., Zuccotti, M., Johnson, K.R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 318-319 (1998).
9. Onishi, A. *et al.* Pig cloning by microinjection of fetal fibroblast nuclei. *Science* **289**, 1188-1190 (2000).
10. Hochedlinger, K. & Plath, K. Epigenetic reprogramming and induced pluripotency. *Development* **136**, 509-523 (2009).
11. Miller, R.A. & Ruddle, F.H. Pluripotent teratocarcinoma-thymocyte somatic cell hybrids. *Cell* **9**, 45-55 (1976).
12. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by *in vitro* hybridization with ES cells. *Current Biology* **11**, 1553-1558 (2001).
13. Taranger, C.K. *et al.* Induction of dedifferentiation, genomewide transcriptional programming, and epigenetic reprogramming by extracts of carcinoma and embryonic stem cells. *Molecular Biology of the Cell* **16**, 5719-5735 (2005).

14. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 1-14 (2006).
15. Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat. Genet.* **41**, 968-976 (2009).
16. Utikal, J., Maherali, N., Kulalert, W. & Hochedlinger, K. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *J Cell Sci*, jcs.054783 (2009).
17. Kim, J.B. *et al.* Direct reprogramming of human neural stem cells by OCT4. *Nature* **461**, 649-643 (2009).
18. Miura, K. *et al.* Variation in the safety of induced pluripotent stem cell lines. *Nature biotechnology* **27**, 743-745 (2009).
19. Park, I.-H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141-147 (2008).
20. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotech* **26**, 795-797 (2008).
21. Chung, T.L. *et al.* Vitamin C promotes widespread yet specific DNA demethylation of the epigenome in human embryonic stem cells. *Stem Cells* **28**, 1848-1855 (2010).
22. Esteban, M.A. *et al.* Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6**, 71-79 (2010).
23. Anokye-Danso, F. *et al.* Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. *Cell Stem Cell* **8**, 376-388 (2011).
24. Wang, Y. *et al.* Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* **40**, 1478-1483 (2008).
25. Wang, Y. & Blelloch, R. Cell cycle regulation by MicroRNAs in embryonic stem cells. *Cancer research* **69**, 4093-4096 (2009).
26. Xu, C. *et al.* Feeder-free growth of undifferentiated human embryonic stem cells. *Nat Biotech* **19**, 971-974 (2001).
27. Miyazaki, T. *et al.* Recombinant human laminin isoforms can support the undifferentiated growth of human embryonic stem cells. *Biochem Biophys Res Commun* **375**, 27-32 (2008).
28. Braam, S.R. *et al.* Recombinant Vitronectin Is a Functionally Defined Substrate That Supports Human Embryonic Stem Cell Self-Renewal via $\alpha V\beta 5$ Integrin. *STEM CELLS* **26**, 2257-2265 (2008).

29. Brafman, D.A. *et al.* LONG-TERM HUMAN PLURIPOTENT STEM CELL SELF-RENEWAL ON SYNTHETIC POLYMER SURFACES. *Biomaterials* **31**, 9135-9144 (2010).
30. Klim, J.R., Li, L., Wrighton, P.J., Piekarczyk, M.S. & Kiessling, L.L. A defined glycosaminoglycan-binding substratum for human pluripotent stem cells. *Nat Meth* **7**, 989-994 (2010).
31. Mei, Y. *et al.* Combinatorial development of biomaterials for clonal growth of human pluripotent stem cells. *Nat Mater* **9**, 768-778 (2010).
32. Villa-Diaz, L.G. *et al.* Synthetic polymer coatings for long-term growth of human embryonic stem cells. *Nat Biotech* **28**, 581-583 (2010).
33. Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T. & Yamanaka, S. Hypoxia enhances the generation of induced pluripotent stem cells. *Cell Stem Cell* **5**, 237-241 (2009).
34. Thorpe, T.A. History of plant tissue culture. *Molecular biotechnology* **37**, 169-180 (2007).
35. Obokata, H. *et al.* Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* **505**, 641-647 (2014).
36. Obokata, H. *et al.* Retraction: Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* **511**, 112 (2014).
37. Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415-419 (2003).
38. Jahner, D. *et al.* De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* **298**, 623-628 (1982).
39. Hotta, A. & Ellis, J. Retroviral vector silencing during iPS cell induction: an epigenetic beacon that signals distinct pluripotent states. *Journal of cellular biochemistry* **105**, 940-948 (2008).
40. Stadtfeld, M., Maherali, N., Breault, D.T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**, 230-240 (2008).
41. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature biotechnology* **26**, 916-924 (2008).
42. Hasegawa, K., Cowan, A.B., Nakatsuji, N. & Suemori, H. Efficient multicistronic expression of a transgene in human embryonic stem cells. *Stem Cells* **25**, 1707-1712 (2007).
43. Cary, L.C. *et al.* Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the

- FP-locus of nuclear polyhedrosis viruses. *Virology* **172**, 156-169 (1989).
44. Gonzalez, F., Boue, S. & Izpisua Belmonte, J.C. Methods for making induced pluripotent stem cells: reprogramming a la carte. *Nature reviews. Genetics* **12**, 231-242 (2011).
 45. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced Pluripotent Stem Cells Generated Without Viral Integration. *Science* **322**, 945-949 (2008).
 46. Li, H.O. *et al.* A cytoplasmic RNA vector derived from nontransmissible Sendai virus with efficient gene transfer and expression. *Journal of virology* **74**, 6564-6569 (2000).
 47. Noemi FUSAKI, H.B., Akiyo NISHIYAMA, Koichi SAEKI, Mamoru HASEGAWA Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy* **85**, 348-362 (2009).
 48. Seki, T. *et al.* Generation of Induced Pluripotent Stem Cells from Human Terminally Differentiated Circulating T Cells. *Cell Stem Cell* **7**, 11-14 (2010).
 49. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors. *Science* **322**, 949-953 (2008).
 50. Gonzalez, F. *et al.* Generation of mouse-induced pluripotent stem cells by transient expression of a single nonviral polycistronic vector. *Proc. Natl. Acad. Sci. USA* **106**, 8918-8922 (2009).
 51. Yu, J. *et al.* Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. *Science* **324**, 797-801 (2009).
 52. Okita, K. *et al.* A more efficient method to generate integration-free human iPS cells. *Nat Meth* **8**, 409-412 (2011).
 53. Chen, Z.Y., He, C.Y., Ehrhardt, A. & Kay, M.A. Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo. *Molecular therapy : the journal of the American Society of Gene Therapy* **8**, 495-500 (2003).
 54. Chen, Z.Y., He, C.Y. & Kay, M.A. Improved production and purification of minicircle DNA vector free of plasmid bacterial sequences and capable of persistent transgene expression in vivo. *Human gene therapy* **16**, 126-131 (2005).
 55. Jia, F. *et al.* A nonviral minicircle vector for deriving human iPS cells. *Nat Meth* **7**, 197-199 (2010).
 56. Warren, L. *et al.* Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. *Cell Stem Cell* **7**, 618-630 (2010).

57. Michiue, H. *et al.* The NH2 terminus of influenza virus hemagglutinin-2 subunit peptides enhances the antitumor potency of polyarginine-mediated p53 protein transduction. *The Journal of biological chemistry* **280**, 8285-8289 (2005).
58. Wadia, J.S. & Dowdy, S.F. Protein transduction technology. *Current opinion in biotechnology* **13**, 52-56 (2002).
59. Keirstead, H.S. *et al.* Human embryonic stem cell-derived oligodendrocyte progenitor cell transplants remyelinate and restore locomotion after spinal cord injury. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **25**, 4694-4705 (2005).
60. Lamba, D.A., Gust, J. & Reh, T.A. Transplantation of human embryonic stem cell-derived photoreceptors restores some visual function in Crx-deficient mice. *Cell Stem Cell* **4**, 73-79 (2009).
61. Yang, D., Zhang, Z.J., Oldenburg, M., Ayala, M. & Zhang, S.C. Human embryonic stem cell-derived dopaminergic neurons reverse functional deficit in parkinsonian rats. *Stem Cells* **26**, 55-63 (2008).
62. Hanna, J. *et al.* Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920-1923 (2007).
63. Chen, Y.F. *et al.* Rapid generation of mature hepatocyte-like cells from human induced pluripotent stem cells by an efficient three-step protocol. *Hepatology* **55**, 1193-1203 (2012).
64. Tucker, B.A. *et al.* Transplantation of adult mouse iPS cell-derived photoreceptor precursors restores retinal structure and function in degenerative mice. *PLoS One* **6**, e18992 (2011).
65. Jeon, K. *et al.* Differentiation and transplantation of functional pancreatic beta cells generated from induced pluripotent stem cells derived from a type 1 diabetes mouse model. *Stem cells and development* **21**, 2642-2655 (2012).
66. Kawamura, M. *et al.* Feasibility, safety, and therapeutic efficacy of human induced pluripotent stem cell-derived cardiomyocyte sheets in a porcine ischemic cardiomyopathy model. *Circulation* **126**, S29-37 (2012).
67. Nori, S. *et al.* Grafted human-induced pluripotent stem-cell-derived neurospheres promote motor functional recovery after spinal cord injury in mice. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16825-16830 (2011).

68. Schwartz, S.D. *et al.* Embryonic stem cell trials for macular degeneration: a preliminary report. *The Lancet* **379**, 713-720 (2012).
69. Robinton, D.A. & Daley, G.Q. The promise of induced pluripotent stem cells in research and therapy. *Nature* **481**, 295-305 (2012).
70. Ebert, A.D. *et al.* Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **457**, 277-280 (2009).
71. Lu, X. & Zhao, T. Clinical therapy using iPSCs: hopes and challenges. *Genomics, proteomics & bioinformatics* **11**, 294-298 (2013).
72. Liu, G.-H. *et al.* Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. *Nature* **472**, 221-225 (2011).
73. Zhang, J. *et al.* A Human iPSC Model of Hutchinson Gilford Progeria Reveals Vascular Smooth Muscle and Mesenchymal Stem Cell Defects. *Cell Stem Cell* **8**, 31-45 (2011).
74. Liu, G.-H. *et al.* Progressive degeneration of human neural stem cells caused by pathogenic LRRK2. *Nature* **491**, 603-607 (2012).
75. Wu, S.M. & Hochedlinger, K. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nature cell biology* **13**, 497-505 (2011).
76. Schwartz, R.E. *et al.* Modeling hepatitis C virus infection using human induced pluripotent stem cells. *Proceedings of the National Academy of Sciences* **109**, 2544-2548 (2012).
77. Trounson, A., Shepard, K.A. & DeWitt, N.D. Human disease modeling with induced pluripotent stem cells. *Current Opinion in Genetics & Development* **22**, 509-516 (2012).
78. Inoue, H., Nagata, N., Kurokawa, H. & Yamanaka, S. iPS cells: a game changer for future medicine. *The EMBO journal* **33**, 409-417 (2014).
79. Mercola, M., Colas, A. & Willems, E. Induced pluripotent stem cells in cardiovascular drug discovery. *Circulation research* **112**, 534-548 (2013).
80. Zhao, X.-y. *et al.* iPS cells produce viable mice through tetraploid complementation. *Nature* **461**, 86-90 (2009).
81. Mikkelsen, T.A. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49-56 (2008).
82. Chin, M.H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111-123 (2009).

83. Ghosh, Z. *et al.* Persistent Donor Cell Gene Expression among Human Induced Pluripotent Stem Cells Contributes to Differences with Human Embryonic Stem Cells. *PLoS ONE* **5**, e8975 (2010).
84. Marchetto, M.C. *et al.* Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS One* **4**, e7076 (2009).
85. Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature biotechnology* **27**, 353-360 (2009).
86. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41**, 1350-1353 (2009).
87. Kim, K. *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat Biotech* **29**, 1117-1119 (2011).
88. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68-73 (2011).
89. Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature cell biology* **13**, 541-549 (2011).
90. Guenther, M.G. *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* **7**, 249-257 (2010).
91. Newman, A.M. & Cooper, J.B. Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell* **7**, 258-262 (2010).
92. Yamanaka, S. Induced pluripotent stem cells: past, present, and future. *Cell Stem Cell* **10**, 678-684 (2012).
93. Polo, J.M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature biotechnology* **28**, 848-855 (2010).
94. Soldner, F. *et al.* Parkinson's Disease Patient-Derived Induced Pluripotent Stem Cells Free of Viral Reprogramming Factors. *Cell* **136**, 964-977 (2009).
95. Hanna, J.H., Saha, K. & Jaenisch, R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143**, 508-525 (2010).
96. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926 (1992).

97. Takahashi, K., Mitsui, K. & Yamanaka, S. Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature* **423**, 541-545 (2003).
98. Varela, E., Schneider, R.P., Ortega, S. & Blasco, M.A. Different telomere-length dynamics at the inner cell mass versus established embryonic stem (ES) cells. *Proceedings of the National Academy of Sciences* **108**, 15207-15212 (2011).
99. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
100. Baker, D.E.C. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nat Biotech* **25**, 207-215 (2007).
101. Mayshar, Y. *et al.* Identification and Classification of Chromosomal Aberrations in Human Induced Pluripotent Stem Cells. *Cell Stem Cell* **7**, 521-531 (2010).
102. Sareen, D. *et al.* Chromosome 7 and 19 Trisomy in Cultured Human Neural Progenitor Cells. *PLoS ONE* **4**, e7630 (2009).
103. Ueyama, H. *et al.* Chromosomal variability of human mesenchymal stem cells cultured under hypoxic conditions. *Journal of Cellular and Molecular Medicine* **16**, 72-82 (2012).
104. Stephenson, E. *et al.* *Safety paradigm: genetic evaluation of therapeutic grade human embryonic stem cells*, Vol. 7. (2010).
105. Hovatta, O. *et al.* A Teratocarcinoma-Like Human Embryonic Stem Cell (hESC) Line and Four hESC Lines Reveal Potentially Oncogenic Genomic Changes. *PLoS ONE* **5**, e10263 (2010).
106. Maitra, A. *et al.* Genomic alterations in cultured human embryonic stem cells. *Nat Genet* **37**, 1099-1103 (2005).
107. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63-67 (2011).
108. Hussein, S.M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58-62 (2011).
109. Zhao, T., Zhang, Z.-N., Rong, Z. & Xu, Y. Immunogenicity of induced pluripotent stem cells. *Nature* **474**, 212-215 (2011).
110. Cheng, L. *et al.* Low Incidence of DNA Sequence Variation in Human Induced Pluripotent Stem Cells Generated by Nonintegrating Plasmid Expression. *Cell Stem Cell* **10**, 337-344 (2012).
111. Young, Margaret A. *et al.* Background Mutations in Parental Cells Account for Most of the Genetic Heterogeneity of Induced Pluripotent Stem Cells. *Cell Stem Cell* **10**, 570-582 (2012).

112. Amariglio, N. *et al.* Donor-Derived Brain Tumor Following Neural Stem Cell Transplantation in an Ataxia Telangiectasia Patient. *PLoS Med* **6**, e1000029 (2009).
113. Banito, A. & Gil, J. Induced pluripotent stem cells and senescence: learning the biology to improve the technology. *EMBO reports* **11**, 353-359 (2010).
114. Guha, P., Morgan, J.W., Mostoslavsky, G., Rodrigues, N.P. & Boyd, A.S. Lack of immune response to differentiated cells derived from syngeneic induced pluripotent stem cells. *Cell Stem Cell* **12**, 407-412 (2013).
115. Goldring, C.E. *et al.* Assessing the safety of stem cell therapeutics. *Cell Stem Cell* **8**, 618-628 (2011).
116. Li, W., Li, K., Wei, W. & Ding, S. Chemical approaches to stem cell biology and therapeutics. *Cell Stem Cell* **13**, 270-283 (2013).
117. Li, W. & Ding, S. Small molecules that modulate embryonic stem cell fate and somatic cell reprogramming. *Trends in pharmacological sciences* **31**, 36-45 (2010).
118. Chambers, S.M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature biotechnology* **27**, 275-280 (2009).
119. Longmire, T.A. *et al.* Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. *Cell Stem Cell* **10**, 398-411 (2012).
120. Hynes, R.O. The Extracellular Matrix: Not Just Pretty Fibrils. *Science* **326**, 1216-1219 (2009).
121. Adamo, L. *et al.* Biomechanical forces promote embryonic haematopoiesis. *Nature* **459**, 1131-1135 (2009).
122. Levenberg, S. *et al.* Differentiation of human embryonic stem cells on three-dimensional polymer scaffolds. *Proceedings of the National Academy of Sciences* **100**, 12741-12746 (2003).
123. Wozniak, M.A. & Chen, C.S. Mechanotransduction in development: a growing role for contractility. *Nat Rev Mol Cell Biol* **10**, 34-43 (2009).
124. Davidoff, A.M. *et al.* Purification of recombinant adeno-associated virus type 8 vectors by ion exchange chromatography generates clinical grade vector stock. *Journal of virological methods* **121**, 209-215 (2004).
125. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-515 (2010).

126. Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols* **8**, 1765-1786 (2013).
127. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139-140 (2010).
128. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols* **7**, 562-578 (2012).
129. Garber, M., Grabherr, M.G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods* **8**, 469-477 (2011).
130. Turro, E., Astle, W.J. & Tavare, S. Flexible analysis of RNA-seq data using mixed effects models. *Bioinformatics (Oxford, England)* **30**, 180-188 (2014).
131. Thomson, J.A. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282**, 1145-1147 (1998).
132. Lerou, P.H. & Daley, G.Q. Therapeutic potential of embryonic stem cells. *Blood Reviews* **19**, 321-331 (2005).
133. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 1-12 (2007).
134. Skottman, H. *et al.* Gene expression signatures of seven individual human embryonic stem cell lines. *Stem Cells* **23**, 1343-1356 (2005).
135. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* **318**, 1917-1920 (2007).
136. Shi, Y. *et al.* A Combined Chemical and Genetic Approach for the Generation of Induced Pluripotent Stem Cells. *Cell Stem Cell* **2**, 525-528 (2008).
137. Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature epub ahead of print* **1 March 2009** (2009).
138. Zhou, H. *et al.* Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins. *Cell Stem Cell* **4**, 381-384 (2009).
139. Kay, M.A. *et al.* Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. *Nat Genet* **24**, 257-261 (2000).
140. Nathwani, A.C. *et al.* Adenovirus-Associated Virus Vector-Mediated Gene Transfer in Hemophilia B. *New England Journal of Medicine* **365**, 2357-2365 (2011).

141. Liu, X. *et al.* Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. *Nature cell biology* **15**, 829-838 (2013).
142. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63 (2008).
143. Lee, R.C., Feinbaum, R.L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-854 (1993).
144. Ruvkun, G. Glimpses of a Tiny RNA World. *Science* **294**, 797-799 (2001).
145. Judson, R.L., Babiarz, J.E., Venere, M. & Blelloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotech* **27**, 459-461 (2009).
146. Houbaviy, H.B., Murray, M.F. & Sharp, P.A. Embryonic Stem Cell-Specific MicroRNAs. *Developmental Cell* **5**, 351-358 (2003).
147. Greer Card, D.A. *et al.* Oct4/Sox2-Regulated miR-302 Targets Cyclin D1 in Human Embryonic Stem Cells. *Molecular and Cellular Biology* **28**, 6426-6438 (2008).
148. Suh, M.-R. *et al.* Human embryonic stem cells express a unique set of microRNAs. *Developmental Biology* **270**, 488-498 (2004).
149. Weltner, J., Anisimov, A., Alitalo, K., Otonkoski, T. & Trokovic, R. Induced Pluripotent Stem Cell Clones Reprogrammed via Recombinant Adeno-Associated Virus-Mediated Transduction Contain Integrated Vector Sequences. *Journal of virology* **86**, 4463-4467 (2012).
150. Okita, K., Hong, H., Takahashi, K. & Yamanaka, S. Generation of mouse-induced pluripotent stem cells with plasmid vectors. *Nat. Protocols* **5**, 418-428 (2010).
151. Lin, S.-L. *et al.* Regulation of somatic cell reprogramming through inducible mir-302 expression. *Nucleic Acids Research* **39**, 1054-1065 (2011).
152. Miyoshi, N. *et al.* Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs. *Cell Stem Cell* **8**, 633-638 (2011).
153. Schlaeger, T.M. *et al.* A comparison of non-integrating reprogramming methods. *Nature biotechnology* **33**, 58-63 (2015).
154. Mahmoudi, S. & Brunet, A. Aging and reprogramming: a two-way street. *Current opinion in cell biology* **24**, 744-756 (2012).
155. Goh, P.A. *et al.* A Systematic Evaluation of Integration Free Reprogramming Methods for Deriving Clinically Relevant Patient

- Specific Induced Pluripotent Stem (iPS) Cells. *PLoS ONE* **8**, e81622 (2013).
156. Bock, C. *et al.* Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439-452 (2011).
 157. Yakubov, E., Rechavi, G., Rozenblatt, S. & Givol, D. Reprogramming of human fibroblasts to pluripotent stem cells using mRNA of four transcription factors. *Biochemical and Biophysical Research Communications* **394**, 189-193 (2010).
 158. Plews, J.R. *et al.* Activation of Pluripotency Genes in Human Fibroblast Cells by a Novel mRNA Based Approach. *PLoS ONE* **5**, e14397 (2010).
 159. Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27-45 (1985).
 160. Reubinoff, B.E., Pera, M.F., Fong, C.-Y., Trounson, A. & Bongso, A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotech* **18**, 399-404 (2000).
 161. Martin, M.J., Muotri, A., Gage, F. & Varki, A. Human embryonic stem cells express an immunogenic nonhuman sialic acid. *Nature medicine* **11**, 228-232 (2005).
 162. Lepperdinger, G., Brunauer, R., Jamnig, A., Laschober, G. & Kassem, M. Controversial issue: is it safe to employ mesenchymal stem cells in cell-based therapies? *Experimental gerontology* **43**, 1018-1023 (2008).
 163. Rottem, S. & Barile, M.F. Beware of mycoplasmas. *Trends in biotechnology* **11**, 143-151 (1993).
 164. Richards, M., Fong, C.Y., Chan, W.K., Wong, P.C. & Bongso, A. Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells. *Nature biotechnology* **20**, 933-936 (2002).
 165. Ellerström, C. *et al.* Derivation of a Xeno-Free Human Embryonic Stem Cell Line. *STEM CELLS* **24**, 2170-2176 (2006).
 166. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat Meth* **8**, 424-429 (2011).
 167. Amit, M. *et al.* Clonally Derived Human Embryonic Stem Cell Lines Maintain Pluripotency and Proliferative Potential for Prolonged Periods of Culture. *Developmental Biology* **227**, 271-278 (2000).

168. Shimizukawa, R. *et al.* Establishment of a new embryonic stem cell line derived from C57BL/6 mouse expressing EGFP ubiquitously. *genesis* **42**, 47-52 (2005).
169. Garcia-Gonzalo, F.R. & Izpisua Belmonte, J.C. Albumin-associated lipids regulate human embryonic stem cell self-renewal. *PLoS One* **3**, e1384 (2008).
170. Sun, N. *et al.* Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells. *Proc. Natl. Acad. Sci. USA* **106**, 15720-15725 (2009).
171. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nature methods* **8**, 424-429 (2011).
172. Park, J.H. *et al.* Establishment and Maintenance of Human Embryonic Stem Cells on STO, a Permanently Growing Cell Line. *Biology of Reproduction* **69**, 2007-2014 (2003).
173. Richards, M. *et al.* Comparative Evaluation of Various Human Feeders for Prolonged Undifferentiated Growth of Human Embryonic Stem Cells. *STEM CELLS* **21**, 546-556 (2003).
174. Xu, C. *et al.* Immortalized Fibroblast-Like Cells Derived from Human Embryonic Stem Cells Support Undifferentiated Cell Growth. *STEM CELLS* **22**, 972-980 (2004).
175. Stojkovic, P. *et al.* An autogeneic feeder cell system that efficiently supports growth of undifferentiated human embryonic stem cells. *Stem Cells* **23**, 306-314 (2005).
176. Takahashi, K., Narita, M., Yokura, M., Ichisaka, T. & Yamanaka, S. Human Induced Pluripotent Stem Cells on Autologous Feeders. *PLoS ONE* **4**, e8067 (2009).
177. Higuchi, A., Ling, Q.D., Ko, Y.A., Chang, Y. & Umezawa, A. Biomaterials for the feeder-free culture of human embryonic stem cells and induced pluripotent stem cells. *Chemical reviews* **111**, 3021-3035 (2011).
178. Xu, C. *et al.* Basic Fibroblast Growth Factor Supports Undifferentiated Human Embryonic Stem Cell Growth Without Conditioned Medium. *STEM CELLS* **23**, 315-323 (2005).
179. Xu, R.-H. *et al.* Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Meth* **2**, 185-190 (2005).
180. Levenstein, M.E. *et al.* Basic Fibroblast Growth Factor Support of Human Embryonic Stem Cell Self-Renewal. *STEM CELLS* **24**, 568-574 (2006).
181. Hughes, C.S., Postovit, L.M. & Lajoie, G.A. Matrigel: a complex protein mixture required for optimal growth of cell culture. *Proteomics* **10**, 1886-1890 (2010).

182. Aumailley, M. *et al.* A simplified laminin nomenclature. *Matrix biology : journal of the International Society for Matrix Biology* **24**, 326-332 (2005).
183. Miner, J.H. & Yurchenco, P.D. Laminin functions in tissue morphogenesis. *Annual review of cell and developmental biology* **20**, 255-284 (2004).
184. Ekblom, P., Lonai, P. & Talts, J.F. Expression and biological role of laminin-1. *Matrix biology : journal of the International Society for Matrix Biology* **22**, 35-47 (2003).
185. Kallunki, P. *et al.* A truncated laminin chain homologous to the B2 chain: structure, spatial expression, and chromosomal assignment. *The Journal of cell biology* **119**, 679-693 (1992).
186. Miner, J.H., Lewis, R.M. & Sanes, J.R. Molecular cloning of a novel laminin chain, alpha 5, and widespread expression in adult mouse tissues. *The Journal of biological chemistry* **270**, 28523-28526 (1995).
187. Melkounian, Z. *et al.* Synthetic peptide-acrylate surfaces for long-term self-renewal and cardiomyocyte differentiation of human embryonic stem cells. *Nat Biotech* **28**, 606-610 (2010).
188. Villa-Diaz, L.G., Ross, A.M., Lahann, J. & Krebsbach, P.H. Concise review: The evolution of human pluripotent stem cell culture: from feeder cells to synthetic coatings. *Stem Cells* **31**, 1-7 (2013).
189. Irwin, E.F., Gupta, R., Dashti, D.C. & Healy, K.E. Engineered Polymer-Media Interfaces for the Long-term Self-renewal of Human Embryonic Stem Cells. *Biomaterials* **32**, 6912-6919 (2011).
190. Nandivada, H. *et al.* Fabrication of synthetic polymer coatings and their use in feeder-free culture of human embryonic stem cells. *Nat. Protocols* **6**, 1037-1043 (2011).
191. Ross, A.M., Nandivada, H., Ryan, A.L. & Lahann, J. Synthetic substrates for long-term stem cell culture. *Polymer* **53**, 2533-2539 (2012).
192. Andrews, P.W. *et al.* Harmonizing standards for producing clinical-grade therapies from pluripotent stem cells. *Nat Biotech* **32**, 724-726 (2014).
193. Harding, J. & Mirochnitchenko, O. Preclinical studies for induced pluripotent stem cell-based therapeutics. *The Journal of biological chemistry* **289**, 4585-4593 (2014).
194. Yu, J., Chau, K.F., Vodyanik, M.A., Jiang, J. & Jiang, Y. Efficient Feeder-Free Episomal Reprogramming with Small Molecules. *PLoS ONE* **6**, e17557 (2011).

195. Mali, P. *et al.* Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells* **28**, 713-720 (2010).
196. Vuoristo, S. *et al.* Laminin isoforms in human embryonic stem cells: synthesis, receptor usage and growth support. *Journal of Cellular and Molecular Medicine* **13**, 2622-2633 (2009).
197. Nakagawa, M. *et al.* A novel efficient feeder-free culture system for the derivation of human induced pluripotent stem cells. *Scientific reports* **4**, 3594 (2014).
198. Aboody, K., Capela, A., Niazi, N., Stern, Jeffrey H. & Temple, S. Translating Stem Cell Studies to the Clinic for CNS Repair: Current State of the Art and the Need for a Rosetta Stone. *Neuron* **70**, 597-613 (2011).
199. Laflamme, M.A. & Murry, C.E. Heart regeneration. *Nature* **473**, 326-335 (2011).
200. Melton, D.A. *Using stem cells to study and possibly treat type 1 diabetes*, Vol. 366. (2011).
201. Si-Tayeb, K., Lemaigre, F.P. & Duncan, S.A. Organogenesis and development of the liver. *Dev Cell* **18**, 175-189 (2010).
202. Hughes, R.D., Mitry, R.R. & Dhawan, A. Hepatocyte transplantation for metabolic liver disease: UK experience. *Journal of the Royal Society of Medicine* **98**, 341-345 (2005).
203. Rashid, S.T. *et al.* Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *The Journal of Clinical Investigation* **120**, 3127-3136 (2010).
204. Zhang, S. *et al.* Rescue of ATP7B function in hepatocyte-like cells from Wilson's disease induced pluripotent stem cells using gene therapy or the chaperone drug curcumin. *Human Molecular Genetics* **20**, 3176-3187 (2011).
205. Cayo, M.A. *et al.* JD induced pluripotent stem cell-derived hepatocytes faithfully recapitulate the pathophysiology of familial hypercholesterolemia. *Hepatology* **56**, 2163-2171 (2012).
206. Taub, R. Liver regeneration: from myth to mechanism. *Nat Rev Mol Cell Biol* **5**, 836-847 (2004).
207. Raju, R., Chau, D., Verfaillie, C.M. & Hu, W.S. The road to regenerative liver therapies: the triumphs, trials and tribulations. *Biotechnology advances* **31**, 1085-1093 (2013).
208. Hamazaki, T. *et al.* Hepatic maturation in differentiating embryonic stem cells in vitro. *FEBS Letters* **497**, 15-19 (2001).
209. Choi, D. *et al.* In Vivo Differentiation of Mouse Embryonic Stem Cells Into Hepatocytes. *Cell Transplantation* **11**, 359-368 (2002).

210. Song, Z. *et al.* Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells. *Cell research* **19**, 1233-1242 (2009).
211. Si-Tayeb, K. *et al.* Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* **51**, 297-305 (2010).
212. Hay, D.C. *et al.* Efficient differentiation of hepatocytes from human embryonic stem cells exhibiting markers recapitulating liver development in vivo. *Stem Cells* **26**, 894-902 (2008).
213. Schier, A.F. & Shen, M.M. Nodal signalling in vertebrate development. *Nature* **403**, 385-389 (2000).
214. Shen, M.M. Nodal signaling: developmental roles and regulation. *Development* **134**, 1023-1034 (2007).
215. Zorn, A.M. & Wells, J.M. Molecular Basis of Vertebrate Endoderm Development, in *International Review of Cytology*, Vol. Volume 259. (ed. W.J. Kwang) 49-111 (Academic Press, 2007).
216. Tremblay, K.D. & Zaret, K.S. Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. *Developmental Biology* **280**, 87-99 (2005).
217. Serls, A.E., Doherty, S., Parvatiyar, P., Wells, J.M. & Deutsch, G.H. Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. *Development* **132**, 35-47 (2005).
218. Deutsch, G., Jung, J., Zheng, M., Lora, J. & Zaret, K.S. A bipotential precursor population for pancreas and liver within the embryonic endoderm. *Development* **128**, 871-881 (2001).
219. Dessimoz, J., Opoka, R., Kordich, J.J., Grapin-Botton, A. & Wells, J.M. FGF signaling is necessary for establishing gut tube domains along the anterior-posterior axis in vivo. *Mechanisms of Development* **123**, 42-55 (2006).
220. Kumar, M., Jordan, N., Melton, D. & Grapin-Botton, A. Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. *Developmental Biology* **259**, 109-122 (2003).
221. McLin, V.A., Rankin, S.A. & Zorn, A.M. Repression of Wnt/ β -catenin signaling in the anterior endoderm is essential for liver and pancreas development. *Development* **134**, 2207-2217 (2007).
222. Tiso, N., Filippi, A., Pauls, S., Bortolussi, M. & Argenton, F. BMP signalling regulates anteroposterior endoderm patterning in zebrafish. *Mechanisms of Development* **118**, 29-37 (2002).
223. Gadue, P., Huber, T.L., Paddison, P.J. & Keller, G.M. Wnt and TGF- β signaling are required for the induction of an in vitro

- model of primitive streak formation using embryonic stem cells. *Proceedings of the National Academy of Sciences* **103**, 16806-16811 (2006).
224. Bossard, P. & Zaret, K.S. GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125**, 4909-4917 (1998).
 225. Zaret, K.S. Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. *Nature reviews. Genetics* **9**, 329-340 (2008).
 226. Calmont, A. *et al.* An FGF response pathway that mediates hepatic gene induction in embryonic endoderm cells. *Dev Cell* **11**, 339-348 (2006).
 227. D'Amour, K.A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotech* **23**, 1534-1541 (2005).
 228. Mfopou, J.K. *et al.* Efficient definitive endoderm induction from mouse embryonic stem cell adherent cultures: a rapid screening model for differentiation studies. *Stem cell research* **12**, 166-177 (2014).
 229. Si-Tayeb, K. *et al.* Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* **51**, 297-305 (2010).
 230. Asgari, S. *et al.* Differentiation and transplantation of human induced pluripotent stem cell-derived hepatocyte-like cells. *Stem cell reviews* **9**, 493-504 (2013).
 231. Medine, C.N., Lucendo-Villarin, B., Zhou, W., West, C.C. & Hay, D.C. Robust Generation of Hepatocyte-like Cells from Human Embryonic Stem Cell Populations. e2969 (2011).
 232. Teo, A.K.K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes & development* **25**, 238-250 (2011).
 233. Hay, D.C. *et al.* Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 12301-12306 (2008).
 234. Sullivan, G.J. *et al.* Generation of Functional Human Hepatic Endoderm from Human iPS cells. *Hepatology (Baltimore, Md.)* **51**, 329-335 (2010).
 235. Schmidt, C. *et al.* Scatter factor/hepatocyte growth factor is essential for liver development. *Nature* **373**, 699-702 (1995).

236. DeLaForest, A. *et al.* HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. *Development* **138**, 4143-4153 (2011).
237. Chang, K.-H. *et al.* Diverse hematopoietic potentials of five human embryonic stem cell lines. *Experimental Cell Research* **314**, 2930-2940 (2008).
238. Mikkola, M. *et al.* Distinct differentiation characteristics of individual human embryonic stem cell lines. *BMC Developmental Biology* **6**, 40 (2006).
239. Buchholz, D.E. *et al.* Derivation of Functional Retinal Pigmented Epithelium from Induced Pluripotent Stem Cells. *STEM CELLS* **27**, 2427-2434 (2009).
240. Spear, B.T. Alpha-fetoprotein gene regulation: lessons from transgenic mice. *Seminars in Cancer Biology* **9**, 109-116 (1999).
241. Tilghman, S.M. & Belayew, A. Transcriptional control of the murine albumin/alpha-fetoprotein locus during development. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 5254-5257 (1982).
242. LeCluyse, E.L. Human hepatocyte culture systems for the in vitro evaluation of cytochrome P450 expression and regulation. *European Journal of Pharmaceutical Sciences* **13**, 343-368 (2001).
243. Gerets, H.H.J. *et al.* Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins. *Cell Biol Toxicol* **28**, 69-87 (2012).
244. Mohsin, S. *et al.* Enhanced hepatic differentiation of mesenchymal stem cells after pretreatment with injured liver tissue. *Differentiation; research in biological diversity* **81**, 42-48 (2011).
245. Yan, Y.B. *et al.* Hepatoprotective effect of an immortal human fetal hepatic cell transplantation on CCL(4)-induced acute liver injury in mice. *Transplantation proceedings* **42**, 2782-2785 (2010).
246. Huang, H.L. *et al.* Hepatoprotective effects of baicalein against CCl(4)-induced acute liver injury in mice. *World journal of gastroenterology : WJG* **18**, 6605-6613 (2012).
247. Chang, H.M. *et al.* Improvement of carbon tetrachloride-induced acute hepatic failure by transplantation of induced pluripotent stem cells without reprogramming factor c-Myc. *International journal of molecular sciences* **13**, 3598-3617 (2012).

248. Moslem, M., Valojerdi, M.R., Pournasr, B., Muhammadnejad, A. & Baharvand, H. Therapeutic potential of human induced pluripotent stem cell-derived mesenchymal stem cells in mice with lethal fulminant hepatic failure. *Cell Transplant* **22**, 1785-1799 (2013).
249. Elliott, S.A. & Sánchez Alvarado, A. THE HISTORY AND ENDURING CONTRIBUTIONS OF PLANARIANS TO THE STUDY OF ANIMAL REGENERATION. *Wiley interdisciplinary reviews. Developmental biology* **2**, 301-326 (2013).
250. Hanahan, D. & Weinberg, R.A. The Hallmarks of Cancer. *Cell* **100**, 57-70 (2000).
251. Cho, R.W. & Clarke, M.F. Recent advances in cancer stem cells. *Curr Opin Genet Dev* **18**, 48-53 (2008).
252. Lobo, N.A., Shimono, Y., Qian, D. & Clarke, M.F. The biology of cancer stem cells. *Annual review of cell and developmental biology* **23**, 675-699 (2007).
253. Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-111 (2001).
254. Bonnet, D. & Dick, J.E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature medicine* **3**, 730-737 (1997).
255. Chen, K., Huang, Y.-h. & Chen, J.-l. Understanding and targeting cancer stem cells: therapeutic implications and challenges. *Acta Pharmacol Sin* **34**, 732-740 (2013).
256. Andrews, P.W., Fenderson, B. & Hakomori, S. Human embryonal carcinoma cells and their differentiation in culture. *International journal of andrology* **10**, 95-104 (1987).
257. Matsuda, T. *et al.* Expression of the receptor tyrosine kinase genes, Ror1 and Ror2, during mouse development. *Mechanisms of Development* **105**, 153-156 (2001).
258. Al-Shawi, R., Ashton, S.V., Underwood, C. & Simons, J.P. Expression of the Ror1 and Ror2 receptor tyrosine kinase genes during mouse development. *Development genes and evolution* **211**, 161-171 (2001).
259. Lyashenko, N. *et al.* Mice lacking the orphan receptor ror1 have distinct skeletal abnormalities and are growth retarded. *Developmental Dynamics* **239**, 2266-2277 (2010).
260. Baskar, S. *et al.* Unique cell surface expression of receptor tyrosine kinase ROR1 in human B-cell chronic lymphocytic leukemia. *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 396-404 (2008).

261. Hudecek, M. *et al.* The B-cell tumor-associated antigen ROR1 can be targeted with T cells modified to express a ROR1-specific chimeric antigen receptor. *Blood* **116**, 4532-4541 (2010).
262. Bicocca, V.T. *et al.* Crosstalk between ROR1 and the Pre-B cell receptor promotes survival of t(1;19) acute lymphoblastic leukemia. *Cancer cell* **22**, 656-667 (2012).
263. Zhang, S. *et al.* The onco-embryonic antigen ROR1 is expressed by a variety of human cancers. *The American journal of pathology* **181**, 1903-1910 (2012).
264. Tang, C. *et al.* An antibody against SSEA-5 glycan on human pluripotent stem cells enables removal of teratoma-forming cells. *Nat Biotech* **29**, 829-834 (2011).
265. Daneshmanesh, A.H. *et al.* Ror1, a cell surface receptor tyrosine kinase is expressed in chronic lymphocytic leukemia and may serve as a putative target for therapy. *International journal of cancer. Journal international du cancer* **123**, 1190-1195 (2008).
266. Barna, G. *et al.* ROR1 expression is not a unique marker of CLL. *Hematological oncology* **29**, 17-21 (2011).
267. Daneshmanesh, A.H. *et al.* Orphan receptor tyrosine kinases ROR1 and ROR2 in hematological malignancies. *Leukemia & lymphoma* **54**, 843-850 (2013).
268. Zhang, S. *et al.* ROR1 is expressed in human breast cancer and associated with enhanced tumor-cell growth. *PLoS One* **7**, e31127 (2012).
269. Borcherdig, N., Kusner, D., Liu, G.H. & Zhang, W. ROR1, an embryonic protein with an emerging role in cancer biology. *Protein & cell* **5**, 496-502 (2014).
270. Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481**, 85-89 (2012).
271. Macfarlane, R.G. An Enzyme Cascade in the Blood Clotting Mechanism, and its Function as a Biochemical Amplifier. *Nature* **202**, 498-499 (1964).
272. Monroe, D.M. & Hoffman, M. The Coagulation Cascade in Cirrhosis. *Clinics in Liver Disease* **13**, 1-9 (2009).
273. Riddel, J.P., Jr., Aouizerat, B.E., Miaskowski, C. & Lillicrap, D.P. Theories of blood coagulation. *Journal of pediatric oncology nursing : official journal of the Association of Pediatric Oncology Nurses* **24**, 123-131 (2007).
274. Hoffman, M. Remodeling the Blood Coagulation Cascade. *J Thromb Thrombolysis* **16**, 17-20 (2003).
275. Kjalke, M. *et al.* Active Site-inactivated Factors VIIa, Xa, and IXa Inhibit Individual Steps in a Cell-based Model of Tissue Factor-

- initiated Coagulation. *Thrombosis and Haemostasis* **80**, 578-584 (1998).
276. Hoffman, M. & Monroe Iii, D.M. A Cell-based Model of Hemostasis. *Thrombosis and Haemostasis* **85**, 958-965 (2001).
 277. Hoffman, M. A cell-based model of coagulation and the role of factor VIIa. *Blood Reviews* **17**, **Supplement 1**, S1-S5 (2003).
 278. Roberts, H.R., Hoffman, M. & Monroe, D.M. A cell-based model of thrombin generation. *Seminars in thrombosis and hemostasis* **32 Suppl 1**, 32-38 (2006).
 279. Ong, K., Horsfall, W., Conway, E.M. & Schuh, A.C. Early embryonic expression of murine coagulation system components. *Thromb Haemost* **84**, 1023-1030 (2000).
 280. Soifer, S.J., Peters, K.G., O'Keefe, J. & Coughlin, S.R. Disparate temporal expression of the prothrombin and thrombin receptor genes during mouse development. *The American journal of pathology* **144**, 60-69 (1994).
 281. Hassan, H.J. *et al.* Blood coagulation factors in human embryonic-fetal development: preferential expression of the FVII/tissue factor pathway. *Blood* **76**, 1158-1164 (1990).
 282. Cui, J., O'Shea, K.S., Purkayastha, A., Saunders, T.L. & Ginsburg, D. Fatal haemorrhage and incomplete block to embryogenesis in mice lacking coagulation factor V. *Nature* **384**, 66-68 (1996).
 283. Rosenberg, R.D. The absence of the blood clotting regulator thrombomodulin causes embryonic lethality in mice before development of a functional cardiovascular system. *Thromb Haemost* **74**, 52-57 (1995).
 284. Bugge, T.H. *et al.* Fatal embryonic bleeding events in mice lacking tissue factor, the cell-associated initiator of blood coagulation. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6258-6263 (1996).
 285. Rosen, E.D. *et al.* Mice lacking factor VII develop normally but suffer fatal perinatal bleeding. *Nature* **390**, 290-294 (1997).
 286. Gasic, G.P., Arenas, C.P., Gasic, T.B. & Gasic, G.J. Coagulation factors X, Xa, and protein S as potent mitogens of cultured aortic smooth muscle cells. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 2317-2320 (1992).
 287. Zhang, Y. *et al.* Thrombomodulin modulates growth of tumor cells independent of its anticoagulant activity. *J Clin Invest* **101**, 1301-1309 (1998).
 288. Cirino, G. *et al.* Factor Xa as an interface between coagulation and inflammation. Molecular mimicry of factor Xa association with

- effector cell protease receptor-1 induces acute inflammation in vivo. *J Clin Invest* **99**, 2446-2451 (1997).
289. Esmon, C.T., Taylor, F.B., Jr. & Snow, T.R. Inflammation and coagulation: linked processes potentially regulated through a common pathway mediated by protein C. *Thromb Haemost* **66**, 160-165 (1991).
 290. Monson, A. *et al.* Factor XIIIa as a nerve-associated transglutaminase. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **12**, 1163-1171 (1998).
 291. Doolittle, R.F. Step-by-step evolution of vertebrate blood coagulation. *Cold Spring Harbor symposia on quantitative biology* **74**, 35-40 (2009).
 292. Fry, H. Blood platelets and coagulation of the blood in marine chordates. 467–503 (1909).
 293. Jiang, Y. & Doolittle, R.F. The evolution of vertebrate blood coagulation as viewed from a comparison of puffer fish and sea squirt genomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7527-7532 (2003).
 294. Doolittle, R.F. The evolution of vertebrate blood coagulation: a case of Yin and Yang. *Thromb Haemost* **70**, 24-28 (1993).
 295. Doolittle, R.F. & Feng, D.F. Reconstructing the evolution of vertebrate blood coagulation from a consideration of the amino acid sequences of clotting proteins. *Cold Spring Harbor symposia on quantitative biology* **52**, 869-874 (1987).
 296. Hughes, A.L. Modes of evolution in the protease and kringle domains of the plasminogen-prothrombin family. *Molecular phylogenetics and evolution* **14**, 469-478 (2000).
 297. Amitrano, L., Guardascione, M.A., Brancaccio, V. & Balzano, A. Coagulation disorders in liver disease. *Seminars in liver disease* **22**, 83-96 (2002).
 298. Dashty, M. *et al.* Characterization of coagulation factor synthesis in nine human primary cell types. *Scientific reports* **2**, 787 (2012).
 299. Andrew, M. The relevance of developmental hemostasis to hemorrhagic disorders of newborns. *Seminars in perinatology* **21**, 70-85 (1997).
 300. Maston, G.A., Evans, S.K. & Green, M.R. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* **7**, 29-59 (2006).
 301. Spitz, F. & Furlong, E.E. Transcription factors: from enhancer binding to developmental control. *Nature reviews. Genetics* **13**, 613-626 (2012).

302. Chan, S.S.-K. & Kyba, M. What is a Master Regulator? *Journal of stem cell research & therapy* **3**, 114 (2013).
303. Zhou, Y. *et al.* TCF7L2 is a master regulator of insulin production and processing. *Hum Mol Genet* **23**, 6419-6431 (2014).
304. Wilhelm, B.T. *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239-1243 (2008).
305. Cloonan, N. *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Meth* **5**, 613-619 (2008).
306. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621-628 (2008).
307. Safdar, H., Inoue, Y., van Puijvelde, G.H., Reitsma, P.H. & van Vlijmen, B.J. The role of hepatocyte nuclear factor 4alpha in regulating mouse hepatic anticoagulation and fibrinolysis gene transcript levels. *Journal of thrombosis and haemostasis : JTH* **8**, 2839-2841 (2010).
308. Safdar, H. *et al.* Modulation of Mouse Coagulation Gene Transcription following Acute *In Vivo* Delivery of Synthetic Small Interfering RNAs Targeting HNF4 α and C/EBP α . *PLoS ONE* **7**, e38104 (2012).
309. Kuo, C.J. *et al.* A transcriptional hierarchy involved in mammalian cell-type specification. *Nature* **355**, 457-461 (1992).
310. Battle, M.A. *et al.* Hepatocyte nuclear factor 4alpha orchestrates expression of cell adhesion proteins during the epithelial transformation of the developing liver. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8419-8424 (2006).
311. Parviz, F. *et al.* Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nat Genet* **34**, 292-296 (2003).
312. Lazarevich, N.L. *et al.* Progression of HCC in mice is associated with a downregulation in the expression of hepatocyte nuclear factors. *Hepatology* **39**, 1038-1047 (2004).
313. Ning, B.F. *et al.* Hepatocyte nuclear factor 4 alpha suppresses the development of hepatocellular carcinoma. *Cancer research* **70**, 7640-7651 (2010).
314. Vogel, C. *et al.* Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Molecular systems biology* **6**, 400 (2010).
315. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343-348 (2011).

316. Hu, J., Li, B. & Kihara, D. Limitations and potentials of current motif discovery algorithms. *Nucleic Acids Res* **33**, 4899-4913 (2005).
317. Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome research* **21**, 447-455 (2011).
318. Yamanaka, S. A fresh look at iPS cells. *Cell* **137**, 13-17 (2009).
319. Watt, F.M. & Driskell, R.R. The therapeutic potential of stem cells. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **365**, 155-163 (2010).
320. Engler, A.J., Sen, S., Sweeney, H.L. & Discher, D.E. Matrix elasticity directs stem cell lineage specification. *Cell* **126**, 677-689 (2006).
321. Evans, N.D. *et al.* Substrate stiffness affects early differentiation events in embryonic stem cells. *European cells & materials* **18**, 1-13; discussion 13-14 (2009).
322. Yim, E.K., Darling, E.M., Kulangara, K., Guilak, F. & Leong, K.W. Nanotopography-induced changes in focal adhesions, cytoskeletal organization, and mechanical properties of human mesenchymal stem cells. *Biomaterials* **31**, 1299-1306 (2010).
323. Wang, P.Y., Yu, J., Lin, J.H. & Tsai, W.B. Modulation of alignment, elongation and contraction of cardiomyocytes through a combination of nanotopography and rigidity of substrates. *Acta biomaterialia* **7**, 3285-3293 (2011).
324. Saha, K. *et al.* Surface-engineered substrates for improved human pluripotent stem cell culture under fully defined conditions. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18714-18719 (2011).
325. Benoit, D.S., Schwartz, M.P., Durney, A.R. & Anseth, K.S. Small functional groups for controlled differentiation of hydrogel-encapsulated human mesenchymal stem cells. *Nat Mater* **7**, 816-823 (2008).
326. Choi, S., Yu, X., Jongpaiboonkit, L., Hollister, S.J. & Murphy, W.L. Inorganic coatings for optimized non-viral transfection of stem cells. *Scientific reports* **3**, 1567 (2013).
327. Zhang, W. *et al.* The synergistic effect of hierarchical micro/nanotopography and bioactive ions for enhanced osseointegration. *Biomaterials* **34**, 3184-3195 (2013).
328. Yang, P.J., Levenston, M.E. & Temenoff, J.S. Modulation of mesenchymal stem cell shape in enzyme-sensitive hydrogels is decoupled from upregulation of fibroblast markers under cyclic tension. *Tissue engineering. Part A* **18**, 2365-2375 (2012).

329. Treiser, M.D. *et al.* Cytoskeleton-based forecasting of stem cell lineage fates. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 610-615 (2010).
330. Hadjipanayi, E., Brown, R.A. & Mudera, V. Interface integration of layered collagen scaffolds with defined matrix stiffness: implications for sheet-based tissue engineering. *Journal of tissue engineering and regenerative medicine* **3**, 230-241 (2009).
331. Saha, K. *et al.* Substrate modulus directs neural stem cell behavior. *Biophysical journal* **95**, 4426-4438 (2008).
332. Nam, J., Johnson, J., Lannutti, J.J. & Agarwal, S. Modulation of embryonic mesenchymal progenitor cell differentiation via control over pure mechanical modulus in electrospun nanofibers. *Acta biomaterialia* **7**, 1516-1524 (2011).
333. D'Amour, K.A. *et al.* Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nature biotechnology* **24**, 1392-1401 (2006).
334. Tibbitt, M.W. & Anseth, K.S. Hydrogels as extracellular matrix mimics for 3D cell culture. *Biotechnology and bioengineering* **103**, 655-663 (2009).
335. Gieni, R.S. & Hendzel, M.J. Mechanotransduction from the ECM to the genome: are the pieces now in place? *Journal of cellular biochemistry* **104**, 1964-1987 (2008).
336. Ferreira, L.S. *et al.* Bioactive hydrogel scaffolds for controllable vascular differentiation of human embryonic stem cells. *Biomaterials* **28**, 2706-2717 (2007).
337. Hwang, N.S. *et al.* In vivo commitment and functional tissue regeneration using human embryonic stem cell-derived mesenchymal cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20641-20646 (2008).
338. Baharvand, H., Hashemi, S.M., Kazemi Ashtiani, S. & Farrokhi, A. Differentiation of human embryonic stem cells into hepatocytes in 2D and 3D culture systems in vitro. *The International journal of developmental biology* **50**, 645-652 (2006).
339. Levenberg, S., Burdick, J.A., Kraehenbuehl, T. & Langer, R. Neurotrophin-induced differentiation of human embryonic stem cells on three-dimensional polymeric scaffolds. *Tissue engineering* **11**, 506-512 (2005).
340. Hwang, N.S., Varghese, S. & Elisseeff, J. Derivation of chondrogenically-committed cells from human embryonic cells for cartilage tissue regeneration. *PLoS One* **3**, e2498 (2008).

341. Ferreira, L. *et al.* Human Embryoid Bodies Containing Nano- and Microparticulate Delivery Vehicles. *Advanced Materials* **20**, 2285-2291 (2008).
342. Spence, J.R. *et al.* Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature* **470**, 105-109 (2011).
343. Lancaster, M.A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373-379 (2013).
344. Nakano, T. *et al.* Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell Stem Cell* **10**, 771-785 (2012).
345. Takebe, T. *et al.* Vascularized and functional human liver from an iPSC-derived organ bud transplant. *Nature* **499**, 481-484 (2013).
346. Shiba, Y. *et al.* Human ES-cell-derived cardiomyocytes electrically couple and suppress arrhythmias in injured hearts. *Nature* **489**, 322-325 (2012).
347. Redmond, D.E., Jr., Vinuela, A., Kordower, J.H. & Isacson, O. Influence of cell preparation and target location on the behavioral recovery after striatal transplantation of fetal dopaminergic neurons in a primate model of Parkinson's disease. *Neurobiology of disease* **29**, 103-116 (2008).
348. Decaens, C., Durand, M., Grosse, B. & Cassio, D. Which in vitro models could be best used to study hepatocyte polarity? *Biology of the cell / under the auspices of the European Cell Biology Organization* **100**, 387-398 (2008).
349. Foty, R.A. & Steinberg, M.S. The differential adhesion hypothesis: a direct evaluation. *Dev Biol* **278**, 255-263 (2005).
350. Ladewig, J., Koch, P. & Brustle, O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol* **14**, 225-236 (2013).
351. Vierbuchen, T. & Wernig, M. Molecular roadblocks for cellular reprogramming. *Molecular cell* **47**, 827-838 (2012).
352. Weintraub, H. *et al.* Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 5434-5438 (1989).
353. Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663-676 (2004).
354. Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375-386 (2010).
355. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035-1041 (2010).

- 356. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D.A. In vivo reprogramming of adult pancreatic exocrine cells to [bgr]-cells. *Nature* **455**, 627-632 (2008).
- 357. Jolma, A. *et al.* DNA-Binding Specificities of Human Transcription Factors. *Cell* **152**, 327-339.
- 358. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nature reviews. Genetics* **10**, 252-263 (2009).